Strategies of Bacterial Gene Expression: Regulatory Mechanisms and Functional Aspects

Georg Fritz

München 2012

Strategies of Bacterial Gene Expression: Regulatory Mechanisms and Functional Aspects

Georg Fritz

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Georg Fritz
aus Gießen

München, den 12. Januar 2012

Erstgutachter: Prof. Dr. Ulrich Gerland
Zweitgutachter: Prof. Dr. Joachim O. Rädler
Tag der mündlichen Prüfung: 15. Februar 2012

Contents

List of Figures

Zusammenfassung

Die Regulation der Genaktivität ist von zentraler Bedeutung für alle lebenden Zellen. Sie ist die Basis für die räumliche und zeitliche Differenzierung von Stammzellen in höheren Organismen und erlaubt es Mikroorganismen schnell auf fluktuierende Wachstumsbedingungen zu reagieren. Auf molekularer Ebene sind diese Reaktionen durch ausgeklügelte Regulationsnetzwerke gesteuert, welche sowohl extrazelluläre als auch intrazelluläre Reize integrieren und dadurch alternative genetische "Programme" in einer sinnvollen, kontext-spezifischen Weise aktivieren. In dieser Dissertation werden drei verschiedene Strategien bakterieller Genregulation untersucht, mit Fokus sowohl auf die molekulare Beschreibung der regulativen Mechanismen als auch auf die funktionalen Aspekte verschiedener Regulationsstrategien.

Um auf schädliche Umwelteinflüsse und Stressfaktoren zu reagieren, besitzen Bakterien zahlreiche Stressantworten, deren schnelle Aktivierung entscheidend für Wachstum und Überleben ist. Die korrekte Funktion vieler Stressantworten hängt jedoch oftmals von einer Kombination verschiedener Umweltfaktoren ab, so wie z.B. von der Verfügbarkeit des Substrats der Stressantwort und von der Abwesenheit inhibitorischer Substanzen. Es ist daher eine zentrale Frage wie Zellen diese Vielzahl von Signalen integrieren, verarbeiten, und so eine konditionale Antwort auf Umwelteinflüsse erzeugen. Der erste Teil dieser Arbeit konzentriert sich auf das Lysin-Decarboxylase-System (Cad) System von *E. coli*, welches ein hervorragendes Beispiel für eine konditionale Stressantwort darstellt. Das Cad-System schützt Zellen gegen Säurestress und wird nur aktiviert wenn sowohl der externe pH-Wert niedrig ist als auch das Substrat für die Stressantwort, Lysin, im Wachstumsmedium vorliegt. Es wird gezeigt, dass das Cad System zusätzlich auch das Endprodukt der Säureantwort, Cadaverin, wahrnimmt, was zu einem verzögerten Ausschalten der Transkription im Cad System führt. Unsere Analyse des molekularen Mechanismus der Cadaverin-Wahrnehmung lässt schliessen, dass der membrangebundene Sensor CadC durch sequentielle Bindung von Cadaverin an vier kooperative Bindungsstellen inaktiviert wird. Solch eine negative Rückkopplung durch Cadaverin könnte als homöostatische "Produce-to-Demand" Strategie dienen, welche die Kosten der Genexpression in Zeiten schwieriger Wachstumsbedingungen minimiert.

Der zweite Teil dieser Dissertation beschäftigt sich mit genetischen Regulationsstrategien mit "Gedächtnis", welche eine wichtige Rolle in der Entwicklung multizellulärer Organismen und in der sozialen Organisation mikrobieller Gemeinschaften spielen. In Analogie zur digitalen Elektronik spricht man bei genetischen Netzwerken mit Gedächtnis von sequentiellen genetischen Schaltkreisen. Während die Prinzipien von rein kombinatorischer Transkriptionslogik schon gut untersucht sind, gibt es bis heute nur wenige Studien zur sequentiellen Transkriptionslogik. In dieser Arbeit werden die Prinzipien der sequentiellen Transkriptionslogik näher untersucht und eine umfassende Klassifikation der elementaren Bausteine für das Design sequentieller Schaltkreise vorgenommen. Unsere Analyse zeigt, dass die einfachsten sequentiellen Bausteine in überraschend wenige funktionale "Familien" fallen. Des Weiteren wird untersucht wie das molekulare Repertoire bakterieller Genregulation verwendet werden kann, um kompakte und robuste Implementationen dieser Bausteine zu erzeugen. Mögliche Anwendungen in der Synthetischen Biologie und die Signifikanz von sequentieller Logik in

natürlichen genetischen Netzwerken werden diskutiert.

Stochastische Fluktuationen kennzeichnen die Expression von Genen auf dem Einzelzell-Niveau und führen selbst in genetisch identischen Zellen unvermeidlich zu heterogenem Verhalten. Während dieses Rauschen oft nachteilige Effekte auf die Funktion einzelner Zellen hat, kann es auch ausgenutzt werden um stochastische Regulationsstrategieen auf Populationsebene zu implementieren. Der letzte Teil dieser Arbeit beschäftigt sich mit der stochastischen Induktion des Ara Systems von *E. coli*, welches den Import und den Abbau des Zuckers Arabinose erlaubt. Es wird gezeigt dass das Ara System nach plötzlicher Zugabe von sub-saturierenden Mengen von Arabinose erst nach einem signifikanten Zeitversatz induziert, welcher zusätzlich eine grosse Zell-zu-Zell-Variation aufweist. Unsere Analyse lässt schliessen, dass dieser heterogene Induktionsprozess mit der stochastischen Verteilung von Arabinose-Aufnahmeproteinen zum Zeitpunkt der Zuckerzugabe korreliert. Der Abschaltvorgang unter Arabinose-Wegnahme ist hingegen schnell und homogen und es wird gezeigt, dass dies von dem vormals uncharakterisierten Membranprotein AraJ abhängt. Ein vereinfachtes mathematisches Modell für bakterielles Wachstum veranschaulicht, dass ein heterogenes Induktionsverhalten grundsätzlich von Vorteil sein kann und in fluktuierenden Umgebungen als "Bet-Hedging" Strategie dienen könnte.

Diese kumulative Dissertation ist folgendermaßen strukturiert: In Kapitel 1 werden die grundsätzlichen Prinzipien von bakterieller Wachstumsphysiologie eingeführt und es wird motiviert warum Genregulation von Vorteil sein kann. In Kapitel 2 wird die quantitative Charakterisierung des Cad Systems präsentiert. Kapitel 3 stellt die Klassifizierung und Charakterisierung sequentieller Schaltkreise in der Genregulation dar. In Kapitel 4 wird das heterogene Anschaltverhalten des Arabinose Systems analysiert und potentielle evolutionäre Vorteile eines solchen Verhaltens untersucht. Die vier Publikationen, welche die Grundlage dieser Dissertation bilden, sind in Anhang A zu finden.

Summary

The regulation of gene expression is of central importance for all living cells. It provides the basis for spatiotemporal differentiation of stem cells in higher organisms and allows for rapid responses of microbes to changing growth conditions. On a molecular level, these responses are controlled by sophisticated regulatory networks, which integrate both environmental as well as intracellular cues to orchestrate alternative gene expression programs in a context-dependent manner. In this thesis, three different strategies of bacterial gene regulation are analyzed, with both a focus on the molecular description of the regulatory mechanisms as well as on the functional aspects of the different regulation strategies.

Under stressful growth conditions the timely activation of specific defense mechanisms can be essential for cellular growth and survival. The proper function of many stress protection systems often relies on a combination of several environmental factors, such as the availability of substrates and the absence of inhibitors of the defense reaction. Therefore, one important challenge is to understand how stress response systems integrate and process multiple input signals and generate responses conditional on the environment. The first part of this thesis focusses on the lysine-decarboxylase (Cad) system of *E. coli*, which is a prime example for a conditional stress response system. The Cad system protects cells against acid stress and is only activated at low pH and if the substrate for the acid defense reaction, lysine, is present in the environment. It is shown that the Cad system also senses the end-product of the defense reaction, cadaverine. This leads to a delayed transcriptional shut-down of the Cad system. Our analysis sheds light on the molecular mechanism of cadaverine sensing, suggesting that the membrane-bound sensor CadC is sequentially inactivated by four cooperatively interacting cadaverine binding sites. Such negative feedback via cadaverine might serve as a homeostatic "produce-to-demand" strategy, which minimizes the cost of operon induction under stressful conditions.

The second part of this thesis addresses gene regulation strategies with memory, which play a vital role in the development of multicellular organisms and in the social organization of microbial communities. In analogy to digital electronics, genetic networks with memory are often referred to as sequential logic circuits. While the schemes of pure combinatorial cis-regulatory logic have been studied in great detail before, little is known about the schemes of sequential logic in gene regulation. Here such a general perspective is provided and a comprehensive classification of the building blocks for sequential logic design is given. Our analysis shows that all basic sequential logic devices fall into surprisingly few functional families, and we scrutinize how the molecular repertoire of bacterial gene regulation can be leveraged to find compact and robust genetic implementations. Possible applications of these new devices in synthetic biology as well as the general significance of sequential logic in natural genetic circuits are discussed.

Stochastic fluctuations are inherent to gene expression at the single cell level and lead to inevitable heterogeneity even in genetically identical cells. While such noise is often detrimental for the proper function of individual cells, it can also be exploited to implement a stochastic regulation strategy at the population level. The last part of this thesis focusses

on the stochastic induction of the Ara system of *E. coli*, which allows cells to import and catabolize the sugar arabinose. Upon sudden addition of intermediate arabinose levels, it is found that individual bacteria induce the Ara system with highly variable time delay. Our analysis suggests that this heterogeneous induction process is correlated with the stochastic distribution of arabinose uptake proteins at the time of sugar addition. In contrast, the down-regulation of the Ara system in response to arabinose removal is rapid and homogeneous and depends on the previously uncharacterized membrane protein AraJ. A mathematical model for bacterial growth reveals that in fluctuating environments the heterogeneous timing in gene induction of the Ara system can naturally arise as a bet-hedging strategy.

This cumulative thesis is structured as follows: In Chapter 1 the basic principles of bacterial growth physiology are introduced, and a motivation why it is beneficial to regulate genes in response to environmental changes is given. In Chapter 2 the quantitative characterization of the Cad system is presented. Chapter 3 addresses the classification and characterization of sequential logic circuits in gene regulatory networks. In Chapter 4 the heterogeneous timing in the induction of the Ara system and its potential functional advantages are analyzed. The full texts of the four publications described within this thesis are attached in Appendix A.

1. Introduction

Microbial cells thrive in complex environments where temporal fluctuations in nutrient levels, toxins and cell density pose challenging growth conditions. The reproductive success of microbes hinges both on their ability to outgrow competitors in times of nutrient availability as well as on their potential to efficiently survive periods of stress and starvation. The adaptation to these ever-changing conditions relies on intricate signaling pathways and gene regulatory networks, which orchestrate both protein production and their activities on all possible levels. It is the aim of this thesis to better understand some of the strategies that bacteria use to control their gene activities – from a mechanistic perspective considering the question of *how* genes are regulated on the one hand, but also from a functional perspective asking *why* genes are expressed in a certain way on the other hand. While the regulatory mechanisms behind a response strategy can be often be deduced from the cellular dynamics upon changing growth conditions, the functional question is typically much harder to answer.

In recent years many experimental and theoretical efforts focussed on the characterization of the evolutionary forces that shape gene regulatory responses in changing environments [39, 63, 82, 100, 105, 141, 158, 167, 171, 197]. For a bacterial cell, the rate of cell proliferation (the "growth rate") is an important characteristic in the race of competing microorganisms. It thus frequently serves as an indicator for the fitness of a cellular phenotype [39, 105]. Therefore, quantifying the dependence of growth rate on the relevant phenotypic parameters is of central importance to the field [141]. In the following, the basic physiology of bacterial growth is reviewed, culminating in simple bacterial 'growth laws' which make a phenomenological link between gene expression and the growth state of a cell [157, 158] (Section 1.1). Here, a trade-off between the costs and benefits associated with enzyme production naturally leads to optimal gene expression levels which maximize the growth rate. However, with temporally varying environmental conditions also the optimal gene expression levels change. Therefore, the last section discusses some of the regulation strategies cells use to adapt their protein contents accordingly (Section 1.2).

1.1. Bacterial Growth Laws in a Nutshell

In balanced exponential growth all constituents of the cell need to be doubled exactly at the same rate [125], involving the reproduction of proteins, polysaccharides, RNA, DNA, and lipids. In a seminal work by Schaechter et al. [154] it was shown that during balanced growth the cellular composition of these macromolecules is mainly a function of the growth rate alone and does not depend on the chemical details of the growth medium. Remarkably, when the growth rate is modulated through the addition of saturating amounts of nutrients of varying quality, the RNA/protein is a linearly increasing function of the growth rate [154] (Fig. 1.1A; *solid line*). Since most of the cellular RNA constitutes ribosomal RNA (85% in *E. coli*), the RNA/protein ratio is proportional to the mass fraction of ribosomes inside a cell [21]. The linearity can then be understood as a consequence of mass balance and under the assumption that ribosomes are limiting during exponential growth [44, 186]. Then the rate

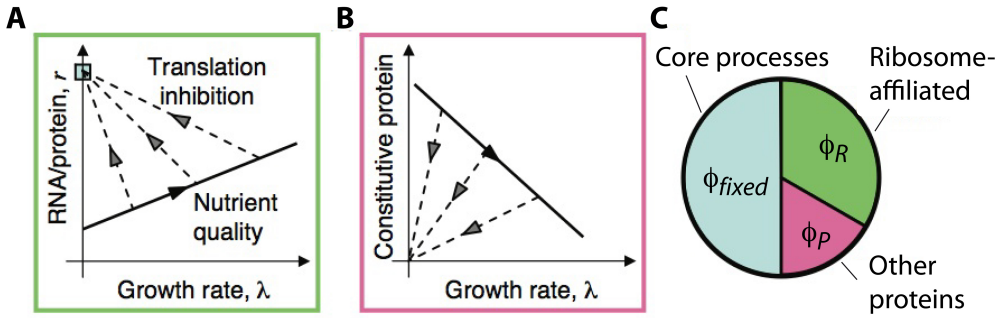


Figure 1.1.: Bacterial growth laws.[1] (**A**) The RNA/protein ratio r (proportional to the ribosomal content) is a linear function of the growth rate, when varied though changes in nutrient quality (*solid line*). Contrary, the RNA/protein ratio is inversely related to growth rate when varied through changes in translational efficiency (*dashed lines*). The maximum value of r under full translational inhibition (*square*) corresponds to a ribosomal mass fraction of 55%, suggesting that there exists a fixed core fraction of the proteome which is unaffected by translational inhibition [158]. (**B**) Mirror-symmetric linear relations are found in the mass fraction of constitutively expressed proteins, leading Scott et al. [158] to a three-component partition model of the proteome in (**C**). It consists of a fixed fraction required for core cellular processes (*blue*), a flexible fraction comprising all ribosome-affiliated proteins (*green*) and another fraction that contains all the remaining proteins, including constitutively expressed ones (*pink*).

of protein mass production is directly proportional to the number of translating ribosomes. Noting that proteins make about 50% of the bacterial dry weight [150], this implies also a linear relation between the RNA/protein ratio and the cell doubling rate [158]. Here the slope $(1/\kappa_t)$ is proportional to the inverse of the translational elongation rate, as has been shown for a family of mutants with altered translation rates [158]. In contrast, the addition of sublethal doses of antibiotics (targeting ribosome function) leads to a linear increase of the RNA/protein ratio with decreasing growth rate (Fig. 1.1; *dashed lines*), with a slope $(1/\kappa_n)$ proportional to the inverse of the nutrient quality.

Another remarkable finding was that the mass fractions of various constitutively expressed proteins without specific regulation display very similar linear relations upon both nutrient- as well as translation-limited growth [158] (Fig. 1.1B). However, the sign of the relations is reversed: with increasing nutrient quality the mass fraction of constitutively expressed protein decreases linearly (Fig. 1.1B; *solid line*), and similarly, translational inhibition leads to a decrease of constitutive protein expression (Fig. 1.1B; *dashed lines*). This mirror-symmetry between the growth-dependencies of ribosomal and constitutive proteins lead Scott et al. [158] to postulate, that there exists a linear constraint between the two fractions. Moreover, from a quantitative analysis of the relations in Figs. 1.1A and B they concluded that the proteome can be divided into three components, including a growth rate-invariant fraction ϕ_{fixed} required for core cellular processes, a fraction ϕ_R of ribosomal proteins and third fraction ϕ_P containing the remainder of the proteome, including constitutive proteins and proteins required for nutrient uptake and processing (Fig. 1.1C). According to their theory the "burden" of unnecessary protein expression [167] resides in adding a fourth fraction to the proteome, which reduces the cellular investment in ribosome synthesis and nutrient allocation and thereby reduces the

[1] Reprinted from Current Opinion in Biotechnology, Volume 22, Matt Scott and Terence Hwa, Bacterial growth laws and their applications, Pages 559-565, Copyright (2011), with permission from Elsevier.

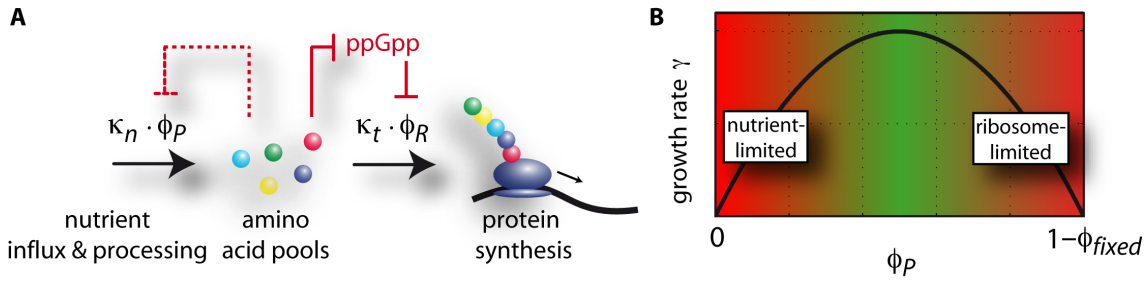


Figure 1.2.: Basic tradeoff in phenomenological theory for bacterial growth.[1] (**A**) The model for bacterial growth assumes that the rate of protein synthesis is proportional to the growth rate of a cell, and that cells can only carry a limited concentration of total protein in their cytoplasm. To optimize the rate of protein synthesis during steady-state exponential growth, the rate of amino acid consumption ($\kappa_t \cdot \phi_R$) must then match the rate of nutrient uptake and processing ($\kappa_n \cdot \phi_P$). This can be achieved by regulating the fractions ϕ_R and ϕ_P through the stringent response (*red lines*), which increases the levels of the alarmone ppGpp upon amino acid shortage and thereby inhibits the transcription of further ribosomes. In addition also nutrient uptake and processing can be regulated directly, e.g., through inhibiting amino acid transporters (*red dashed line*). (**B**) Qualitative illustration of the growth rate as a function of varying the fraction ϕ_P of proteins required for nutrient allocation. For low ϕ_P growth is limited by nutrient uptake and ribosomes do not operate at their full rate, whereas at high ϕ_P ribosomes become limiting, such that amino acids accumulate in the cytoplasm and are transported in and out of the cell in futile, energy-demanding cycles.

growth rate [157].

Model for Bacterial Growth

Given that ϕ_R and ϕ_P can be flexibly interchanged under the constraint that their sum adds up to a constant,

$$\phi_R + \phi_P = 1 - \phi_{fixed}\,, \tag{1.1}$$

what determines the optimal balance between the two fractions under a given growth condition? To approach this problem, Scott et al. [158] proposed a simple phenomenological theory for bacterial growth (Fig. 1.2A). In their model, the ribosomal fraction ϕ_R exerts a positive effect on the growth rate γ through increasing the rate of protein mass accumulation ($\gamma \propto \phi_R$). Likewise, the fraction ϕ_P has also a positive effect on growth, since it is proportional to the rate of nutrient import and processing, thereby increasing the flux of amino acids into the cell ($\gamma \propto \phi_P$). Taken into account the linear constraint in Eq. (1.1), the growth rate then qualitatively displays a maximum at intermediate values of, say, ϕ_P (Fig. 1.2B). Only if the rate of amino acid consumption by protein synthesis exactly balances the rate of nutrient influx into the cell, resources can be optimally allocated. Then the growth rate is maximized and equal to both fluxes, i.e.,

$$\gamma = \kappa_n \cdot \phi_P \tag{1.2}$$
$$\gamma = \kappa_t \cdot \phi_R\,. \tag{1.3}$$

[1] Figure, part **A** adapted from Scott et al. [158].

If ϕ_P is too small, the growth rate is limited by nutrient influx and ribosomes cannot proceed at their maximal speed, as illustrated in Fig. 1.2B. Likewise, if ϕ_P is too large, ribosomes become limiting and intracellular nutrient pools accumulate, leading to futile cycling of excess amino acids in and out of the cell.

On a molecular level, the ribosome partition (ϕ_R, ϕ_P) can be coordinated by the so-called stringent response (Fig. 1.2A; *red lines*). Upon amino acid shortage the number of uncharged tRNAs increases and stimulates synthesis of the alarmone ppGpp. The increase of ppGpp results in transcriptional inhibition of ribosomal promoters and thereby in a reduction of amino acid outflux. While changes in ϕ_R indirectly regulate also the influx of amino acids via changes in ϕ_P, there are also direct mechanisms targeting, e.g., the production of specific amino acids transporters or biosynthetic pathways (Fig. 1.2A; *red dashed line*).

In Eqs. (1.1)-(1.3), the growth rate and the ribosome partition (ϕ_R, ϕ_P) are dictated by the environmental and cellular parameters κ_n, κ_t, and ϕ_{fixed}. Solving, e.g., for the growth rate yields

$$\gamma = \lambda_{max} \cdot \frac{\kappa_n}{\kappa_n + \kappa_t}\,, \tag{1.4}$$

with $\lambda_{max} = (1 - \phi_{fixed}) \cdot \kappa_t$ being the maximally attainable growth rate. Notably, within this "growth law" the growth rate follows a Michaelis-Menten form as a function of the substrate *quality* κ_n. This expression is analogous to the famous Monod-relation [125], which describes the growth rate under nutrient-limited growth in a chemostat, with the substrate *concentration* $[S]$ taking the place of κ_n. Similar expressions to (1.4) can be obtained for ϕ_R and ϕ_P, consistently reproducing the empirical findings on the linear relationships shown in Figs. 1.1A and B. Together, the phenomenological theory for bacterial growth formulated by Scott et al. [158] provides useful links between many important but seemingly unrelated aspects of bacterial physiology and helps to better understand and control the function of endogenous and synthetic circuits.

Optimal Gene Expression Levels

The previous section showed that during balanced exponential growth, a limited protein carrying capacity leads to a trade-off between ribosomes and proteins required for nutrient allocation. The origin of this limitation likely arise from molecular crowding effects [170], where an increase in total protein concentration decreases the diffusion constant and thereby reduces the enzymatic activity of all proteins within the cell. Likewise, under different growth conditions also other physical constraints can limit gene expression and give rise to optimal expression levels. For instance, during stationary phase the transcription of one set of genes can reduce the level of transcription of other genes, suggesting that sigma factors compete for a limiting amount of RNA polymerase [50]. Similarly, membrane space required to accommodate essential nutrient transporters, sensors and alike, represents another potential bottleneck [140]. In fact, the membrane is densely packed with about 200,000 proteins (60%vol protein; 40%vol lipid) [169] and overexpressed membrane protein affects integrity of the membrane and thus cell viability, leading to, e.g., reduced growth and hampered division [187]. Taken together, trade-offs arise ubiquitously within microbial cells, all sharing an effective "cost" and an effective "benefit" of protein production. As a result, cellular growth is maximized at intermediate enzyme expression levels.

1.2. Strategies of Bacterial Gene Regulation

Typically, the optimal expression level of a protein strongly depends on the external growth conditions. In the face of temporal fluctuations of nutrients and/or growth-inhibitory substances within the microenvironment of a cell, it is crucial to precisely monitor these changes and to adapt their gene expression levels accordingly [39, 82, 141, 171]. For instance, if *Escherichia coli* cells grow on a mixture of glucose and lactose, they first completely deplete the preferential carbon source, glucose, from the medium before they initiate synthesis of the enzymes necessary for lactose metabolism [126]. Generally, this multidimensional regulation process is controlled by a web of interdependent signal transduction and gene regulatory networks. They integrate and process external input signals to control all levels of protein expression, ranging, e.g., from signal perception via heterogeneous receptor clusters [121, 165], over signal integration at the level of cis-regulatory transcription control [23, 97, 159] to post-transcriptional regulation of transcript stability via small noncoding RNA (sRNA) [103, 106].

Independent of the mechanism behind the regulatory control, bacteria can use various strategies to respond to environmental changes. For instance, a colony of bacterial cells can respond either in a deterministic or in a stochastic fashion to an external signal. In the former strategy, bacterial cells often use feedback regulation to suppress gene expression noise, leading to a largely uniform behavior within all cells of a culture (pure strategy). Contrary, in the latter strategy, cells often amplify and exploit stochastic fluctuations to generate a large spectrum of phenotypic heterogeneity (mixed strategy). In addition, both strategies can be complemented with memory, meaning that individual cells 'remember' some of the extracellular cues experienced during their ancestral history. In the following, I will introduce the basic ingredients required to implement some of these strategies and briefly discuss their functional advantages.

Combinatorial Gene Expression Strategies

One major focus of this thesis is to understand how cells integrate multiple environmental input signals and process this information to generate an optimal cellular response. Moreover, the induction of such responses frequently affect the levels of the input signals themselves, leading to a rich spectrum of nonlinear feedback interactions accompanied with interesting dynamical behavior. For instance, under certain parameters negative feedback regulation can lead to transient expression kinetics with homeostatic control of target gene expression in the long run [76, 137], or, if the inherent delays of the regulatory cascades exceed a critical value, negative feedback can also give rise to oscillatory dynamics [107, 168]. Likewise, bi- or multistable expression patterns may emerge from positive feedback regulation – as long as it goes along with sufficiently cooperative regulatory control [51, 134]. Hence, the dynamical behavior of genetic and signal processing circuits critically hinges, both qualitatively and quantitatively, on the biochemical details of the regulatory interactions. Therefore, a systems level analysis of the cellular dynamics can reveal important insight into molecular interaction parameters which are often difficult to access experimentally under *in vivo* conditions.

Chapter 2 of this thesis presents such a systems-level study for a specific pH stress response module of *E. coli*, which is conditionally activated only if the conditions for a promising acid defense are given. However, the system is down-regulated after a transient expression phase although the primary stimuli for its induction persists. As will be shown, this behavior emerges from a negative feedback via the end-product of the acid-defense mechanism.

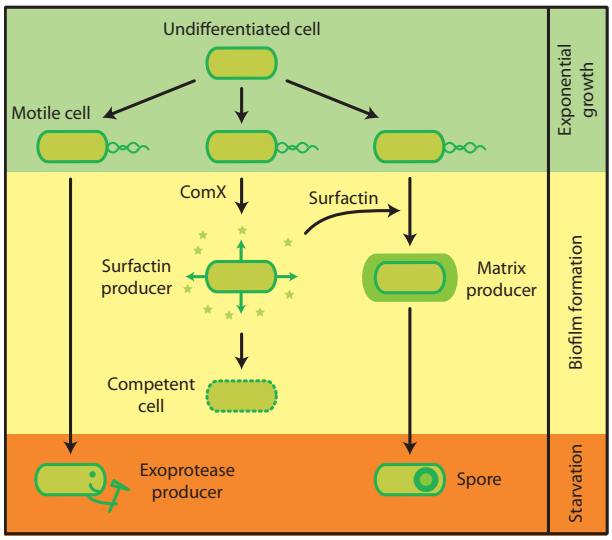


Figure 1.3.: Schematic representation of the cell differentiation process in communities of *Bacillus subtilis* cells.[1]

The microscopic details of this feedback mechanism are resolved through the combination of experimentation and modeling. Since induction of the acid defense module is associated with a large cost under stressful conditions, it seems likely that the transient induction is the result of a homeostatic "produce-to-demand" strategy, in which the negative feedback minimizes the cost of operon induction and saves cellular resources for the investment into other cell-protecting processes.

Microbial Strategies with Memory

While such "responsive switching" between phenotypes is advantageous whenever the cost for maintaining the sensing apparatus is small and if the uncertainty of environmental changes is high [99], cells may also encounter more regular environments in which correlations in growth conditions offer the opportunity to prepare in advance for the next environmental change [123]. A central requirement for such adaptive responses is the bacteria's ability to "memorize" environmental conditions, e.g., in the form of distinct gene expression states, and to inherit those to their progeny. In fact, such epigenetic memory is not only restricted to stem cell differentiation in higher eukaryotes, but is also ubiquitously found in microbial communities [28, 29, 85, 149, 183, 194]. For example, the soil bacterium *Bacillus subtilis* uses environmental cues to differentiate into numerous subpopulations of specialized cell types (Fig. 1.3), which together form a highly structured and competitive biofilm [111].

Despite the prevalence of memory in gene regulatory circuits, a systematic overview over the repertoire of bacterial memory "devices" is still missing to date. In analogy to digital electronics, where circuits with memory are referred to as sequential logic devices, Chapter 3 provides a comprehensive classification of sequential logic in gene regulatory circuits. Moreover, theoretical analysis demonstrates how a combination of feedback regulation and simple protein-protein interactions can be leveraged to implement compact and robust memory devices in a biologically relevant parameter regime. The construction and experimental

[1] Figure adapted in simplified form López and Kolter [111].

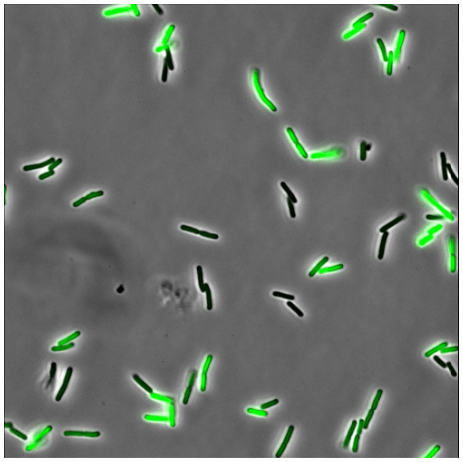


Figure 1.4.: Noisy gene expression in the arabinose system of *E. coli* at intermediate levels of the inducer arabinose.[1]

characterization of these circuits would be a milestone for synthetic biology, and it seems likely that some of the proposed schemes will be identified in naturally occurring genetic networks.

Functional Gene Expression Noise

It has long been known that colonies of genetically identical microbes in homogeneous environments can display substantial cell-to-cell variability in gene expression states [116, 130]. Such phenotypic heterogeneity has been characterized with modern single-cell methods in a range of microbial systems [15, 113, 161]. While noise is omnipresent in gene regulatory circuits [146], such that phenotypic heterogeneity is easy to produce, the intriguing question arises whether the noise is only a necessary evil or also plays a functional role [48, 55, 56, 94, 199]. Generally, one expects phenotypic heterogeneity to be useful as a diversification strategy in risky environments ("bet-hedging") [3, 8, 11, 113, 136] or as a division-of-labor strategy in bacterial communities [6, 112, 182]. A well established example for microbial bet-hedging is bacterial persistence, where a clonal population of *E. coli* copes with unpredictable exposures to antibiotics by stochastically switching between a slow-growing resistant and a fast-growing non-resistant phenotype [8]. However, such a "stochastic switching strategy" for gene regulation is beneficial only for specific systems, which alleviate sudden severe stresses [99, 179].

Most regulatory systems appear to implement a responsive switching strategy, which is advantageous whenever the cost of the investment into sensing and signal transduction is outweighed by the benefit of adaptive gene expression (see above) [99]. So far, stochastic and responsive switching have generally been regarded as competing and mutually exclusive strategies. One might ask, whether there also exist situations in which a regulated stochastic response, i.e., a combination of sensing and stochastic switching, can be a strategy that optimizes growth and survival. Paradigmatic examples of genetic circuits that implement responsive switching are the inducible carbon utilization systems for sugars such as lactose, arabinose or galactose [2, 15, 85, 130, 134]. Under conditions typically realized in the lab, these circuits appear to accomplish a clearcut deterministic strategy, where the proteins for uptake and digestion of the sugar are only produced when needed. However, single-cell

[1] With friendly permission by Dr. Judith Megerle, LMU Munich.

studies have shown that these prototypical regulation systems display significant cell-to-cell heterogeneity in gene expression at low sugar concentrations: For instance, the distribution of *lac* operon expression is relatively broad and even becomes bimodal under specific (artificial) conditions [130, 134]; the *ara* system displays similar behavior [161], see Fig. 1.4. If there is an optimal expression level for a given sugar concentration [39], such heterogeneous behavior seems counterintuitive, as it is expected to reduce the average growth rate [171, 191].

In Chapter 4 it is argued that a full understanding of the regulation of these systems necessarily has to factor in the *timing* of gene induction, which is demonstrated to be broadly heterogeneous in the *ara* operon at intermediate inducer levels. Whilst analyzing the molecular mechanisms behind this "responsive stochastic switching" behavior, the functional question, as to whether responsive stochastic switching confers an advantage over other possible regulation schemes, is addressed. To that end a coarse-grained mathematical model for bacterial growth in unpredictable fluctuating environments will now be devised. An essential property of the model is a transient risk associated with induction of the system, arising from a cost-benefit analysis of gene expression where protein synthesis incurs an immediate cost and only a delayed enzymatic benefit. Given such a finite "amortization time" and an unpredictably fluctuating environment, it is show that the observed responsive stochastic switching naturally arises as a regulated bet-hedging strategy.

2. Conditional pH Stress Response in Escherichia coli

Part of the work described in this chapter was [58] or will be published [72]. The full articles are attached in Appendices A.1 (Paper I) and A.2 (Paper II), respectively.

Rapid adaptation to adverse environmental conditions is key to the reproductive success of microbes. For instance, gastroninternal bacteria, such as *E. coli*, thrive within and outside the human host, where growth conditions often change suddenly and deviate drastically from their optimal ranges: In the outside world microorganisms must be able to adapt rapidly to extreme variations in salinity, temperature or osmolarity, whereas entry into the host is strongly dependent on bacteria's ability to withstand highly acidic conditions in the stomach. Adaptation to these ever-changing conditions requires the accurate monitoring of critical parameters and a precise and specific information flow in order to respond with appropriate alterations in gene expression and protein activity. Therefore, one important challenge is to understand how stress response systems integrate and process multiple input signals and generate responses conditional on the environment.

The work presented in this chapter focusses on the lysine-decarboxylase (Cad) system of *E. coli* and arose in fruitful collaboration with the group of Prof. Dr. Kirsten Jung (Microbiology, LMU Munich). The Cad system is conditionally induced at acidic pH and only in lysine-rich environments and thus serves as a prime example for a conditional stress response module. Interestingly, as will be shown below, the system displays only a transient expression dynamics although the conditions for its induction - low external pH and high lysine abundance - persist. This is unlike other stress responses, where typically the stress is directly counteracted until it returns to physiological levels. As a possible explanation for the transient response in the lysine-decarboxylase system, it was hypothesized that the end-product of the stress response, cadaverine, exerts a negative feedback on Cad expression [128] and thus might serve as a third regulatory input to the Cad system.

Here this hypothesis was tested by a combined experimental and modeling approach, aiming at a comprehensive quantitative description of the Cad module, which can be used to predict the system's output under various experimental conditions (Paper I). In subsequent work we leveraged our model to analyze the *in vivo* dynamics of the Cad module with several mutants of the sensor CadC. In conjunction with structural information on CadC, this allowed for the inference of a mechanistic model of feedback inhibition via cadaverine, according to which CadC is sequentially inactivated by cadaverine by binding to two cooperative binding sites (Paper II). During the course of this thesis, the remarkable ability to integrate three signals by a single polypeptide has been the focus of extensive research, and significant progress towards a molecular understanding of signal integration by the one-component system CadC has been made [47, 58, 71, 173–176]. The results presented in this chapter add on this by shedding light on the mechanisms of cadaverine-dependent repression and thereby advance our understanding of differential gene regulation in bacteria.

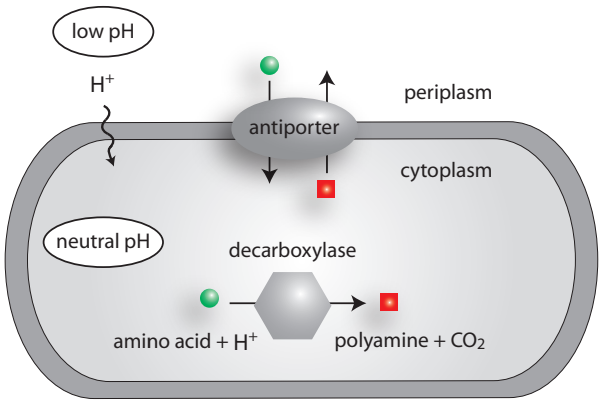


Figure 2.1.: Basic function of the amino acid decarboxylase systems in *E. coli*. At low external pH levels protons (H^+) leak across the plasma membrane and decrease the neutral pH of the cytoplasm. During decarboxylation of amino acids intracellular protons are consumed and the cognate antiporters expel the decarboxylation products (polyamines) in exchange for importing new amino acid substrates, effectively expelling one proton per transport cycle.

I will start out this chapter with a brief introduction into bacterial responses to acidic stress, with a special focus on the amino acid decarboxylase systems (Section 2.1). Then the biochemical properties of the molecular players within the Cad module are introduced, serving as a basis for the formulation of a quantitative mathematical model for the regulatory dynamics of the Cad system (Section 2.2). In order to optimally explain the experimental observations, model parameters – such as reaction rates and equilibrium constants – are calibrated by parameter estimation techniques, and a brief overview of standard methods is given (Section 2.3). At the end of the chapter, the central results of Papers I and II are discussed (Sections 2.4 and 2.5). Finally, I will give an outlook on possible applications of our model in a systems level study of *E. coli*'s acid stress response (Section 2.6).

2.1. Bacterial Responses to Acid Stress

Like many other enteric bacteria, *E. coli* is a neutralophile, that is, it grows best under neutral pH conditions [60]. However, acid stress is its daily business, as, upon ingestion, bacteria are instantly faced with the extreme low pH of the stomach (pH 2-4) and are later exposed to volatile fatty acids (VFA) in the intestines (pH 5-6). Although the cytoplasmic membrane provides an efficient barrier against charged compounds, protons may leak across the membrane at extremely low external pH and acidify the cytoplasm [193]. Likewise, weak acids, such as VFA, can penetrate the membrane already at higher pH levels in their protonated form and lower the internal pH by dissociation in the cytoplasm. This reduction of internal pH reduces the proton motive force [200] and may have deleterious effects on structural integrity of proteins and DNA as well as on many other biochemical processes within the cell [53].

pH Homeostasis and Acid Tolerance Response

To be successful under such harsh conditions, *E. coli* has evolved the remarkable ability to sustain growth at any pH between 4.5 and 9 and keeps its doubling time almost unaffected between pH 5.8 and 8 [60]. This behavior is implemented in a battery of systems that enables cells to keep the internal pH constant (pH homeostasis) or, if these mechanisms fail, to cope with low internal pH (ATR, *acid tolerance response*). While during normal growth the buffering capacity of the cytoplasm provides a certain passive ability of pH homeostasis, cells also use diverse active mechanisms to keep the internal pH constant [96, 162]. These mechanisms include metabolic switching to generate acidic or neutral end-products and proton expulsion via primary proton pumps, H^+-ion transporters, or, as involved in the amino acid decarboxylase systems, amino acid-polyamine antiporters (see below). Likewise, during ATR so-called acid shock proteins are expressed, which shield and repair important cytoplasmic macromolecules and thereby protect bacteria from the lethal effects of low internal pH [10].

The rapid relaxation of internal pH back to physiological levels already a few minutes after acid stress suggests, that some homeostasis systems are constitutively expressed in *E. coli* [193]. However, as seen in the reduced growth rates of acidophiles at neutral pH, this preparedness comes at the cost of expressing unnecessary proteins under conditions where they are not needed [96]. Therefore, gastrointestinal bacteria have evolved sophisticated regulation strategies to assert that most of the costly defense mechanisms are only mounted in response to pH stress. In fact, the full function of the ATR crucially relies on a certain pre-adaptation under mildly acidic conditions (pH 5-6), enabling *E. coli* to survive strongly acidic conditions (∼pH 1.5) for several hours [66]. But also other acid resistance modules are induced only when needed, with a total number of 160 pH-regulated genes [117].

Amino Acid Decarboxylase Systems

Among the best-studied acid defense mechanisms are the inducible amino acid decarboxylase systems, the most important of which comprise the glutamate, arginine, lysine and ornithine decarboxylase systems [53]. While their contribution to pH homeostasis is already well understood (see below), their role in ATR is just beginning to be revealed [80]. Together, they all share a simple mode of action, which is based on the combined activity of a decarboxylase and its cognate antiporter (Fig. 2.1): The decarboxylase enzymes convert their amino acid substrates under consumption of a cytoplasmic proton into a polyamine and CO_2. Subsequently, the antiporter exports the decarboxylation products in exchange for importing new amino acid substrates. The net result of this cycle is the removal of protons from the cytoplasm and the expulsion of basic polyamines into the environment, effectively increasing the pH both inside as well as outside the cell [151]. Besides these primary defenses, periplasmic polyamines reduce the permeability of the outer membrane porin OmpC by 70 % [152] and the associated reduction of acid influx increases the ability to survive pH stress significantly [153]. Yet another aspect is the production of CO_2, which is proposed to allow further growth under acidic conditions where compensating the loss of CO_2 at low pH by chemical equilibrium [19].

2.2. The Cad System of E. coli

The central components of the Cad system are the enzyme CadA, the transport protein CadB, and the regulatory protein CadC, see Fig. 2.2. The decarboxylase CadA converts the amino

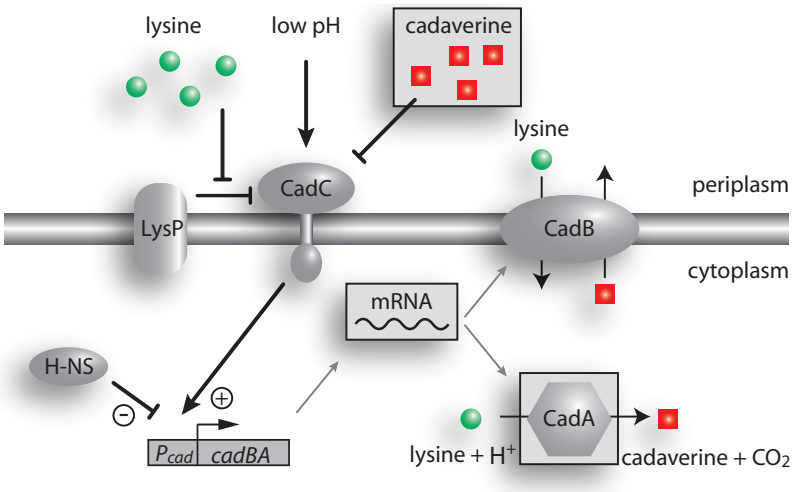


Figure 2.2.: Regulation of the lysine decarboxylase system in *E. coli*. For a detailed description please refer to the main text.

acid lysine into cadaverine [151] and the antiporter CadB imports the substrate, lysine, and exports the product, cadaverine [122, 164]. Together, CadA and CadB reduce the intracellular H^+ concentration and thereby contribute to pH homeostasis, as described above. The cytoplasmic membrane protein CadC not only senses the external conditions [41, 129, 192] but also regulates the response by binding directly to the DNA and activating the transcription of *cadBA* [98]. Similar to other members of the ToxR family [139], CadC thereby performs signal transduction in a single component, without the phosphorylation step employed by two-component systems [101]. Therefore, these proteins transduce information across the lipid bilayer without involving chemical modification and thus represent the simplest known transmembrane signaling systems. However, Fig. 2.2 also depicts the lysine permease LysP, which is not part of the *cad* operon but essential for its function, since CadC senses lysine indirectly via interaction with LysP [41, 129, 175, 192]. In contrast, the external (periplasmic) pH is believed to be sensed directly by CadC through a pH-dependent conformational transition [71] and/or proteolytic cleavage [102].

The signal integration performed by CadC ensures that CadA and CadB are produced only under the combined conditions of low pH and extracellular lysine abundance. However, besides satisfying this physiological requirement, the Cad system also senses a third signal, external cadaverine, which was found to repress the long-term expression of the *cad* operon and might thus constitute a negative feedback in the system [129]. While such a negative feedback would be a plausible explanation for the observation that the expression of *cadBA* is transient [128], similar dynamics could result either from the depletion of external lysine from the medium or by a raise of external pH during the course of these experiments. To discriminate between these three scenarios and to test whether cadaverine is in fact a negative feedback regulator of the Cad module, we established a quantitative mathematical model for the Cad system and calibrated the model against a rich set of experimental data (see Sec. 2.4 below). Before discussing the main results of our study, however, I first introduce the methods used for quantitative data analysis.

2.3. Data Analysis

The analysis of gene expression and signaling networks with the help of quantitative mathematical models is an essential tool to link the behavior of a system to the interactions between its components [81]. This link critically relies on the calibration of model parameters with a suitable experimental data set in a process called data regression. Here the most important methods for parameter estimation and for the derivation of confidence intervals on the parameter estimates are reviewed, closely following the book chapter by Geier et al. [62]. Along these lines, a Monte-Carlo method for global optimization and for the identification of 'sloppy', that is unidentifiable, parameters is presented.

Parameter Estimation

Typical measurements in molecular biology are performed of large samples of bacterial cells, since sufficient amounts of cell extract are required for the precise determination of cellular constituents, such as mRNA or protein levels. Despite the large number of cells used for each preparation, each measurement is subject to inevitable preparation uncertainties, such that the measured data y_{ij} is the sum of the 'true' value of the population-averaged observable $y_i(t_j)$ at time t_j and a measurement error ϵ_{ij},

$$y_{ij} = y_i(t_j) + \epsilon_{ij} \,. \tag{2.1}$$

Here the measurement errors are often assumed to be independent across all measurements and time points and follow Gaussian distributions with zero mean and variance σ_{ij}^2. While this assumption sets the basis for many statistical tests applied in post-regression diagnostics [95], in practice there are often systematic deviations from normality, such that the statistical requirements for the application of these methods are not fulfilled. In such cases, more sophisticated, non-parametric Monte-Carlo methods without specific error-models can be applied [95].

If one neglects fluctuations in the biochemical reaction network, the dynamical behavior of the average molecule numbers of each chemical species can be described by a set of ordinary differential equations. If we let $\mathbf{x}(t)$ be the vector of the N state variables at time t, $\mathbf{k}$ be the vector of all parameters and $\mathbf{x}_0$ the vector of initial conditions, the set of reaction rate equations can be expressed as

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{k}) \,, \tag{2.2}$$

where $\mathbf{f}(\cdot)$ is a continuously differentiable function with respect to its parameters. As most commonly not all state variables can be monitored experimentally and frequently only combinations of them are assessed, an additional observation function relates the state variables to the M dimensional vector of observables $\mathbf{y}(t)$,

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \mathbf{s}) \,. \tag{2.3}$$

where $\mathbf{g}(\cdot)$ involves additional parameters $\mathbf{s}$, which are required to scale species concentrations in the model to experimentally measurable quantities, such as protein activities or relative mRNA levels. Together with the initial states the set of parameters $\theta = \{\mathbf{k}, \mathbf{s}, \mathbf{x}_0\}$ fully specifies the model.

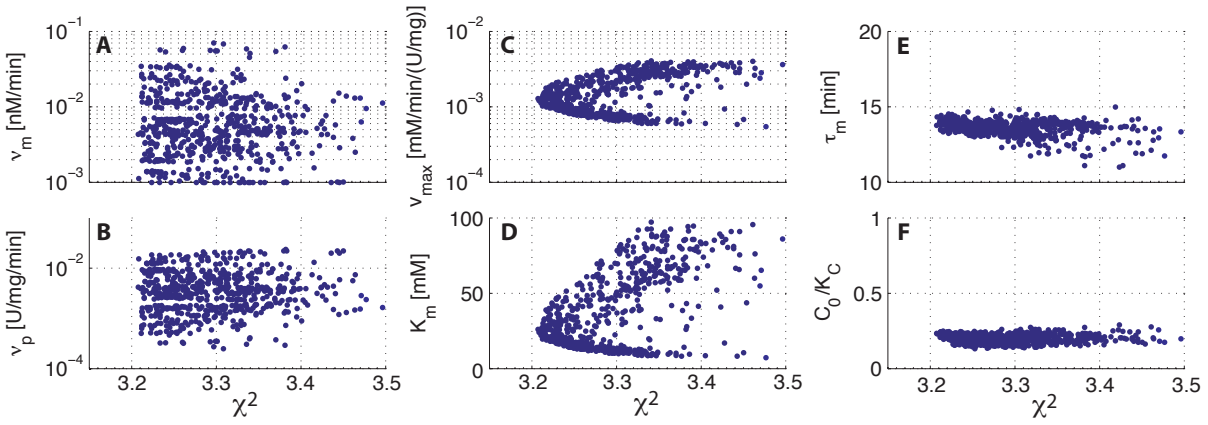


Figure 2.3.: Correlations between the goodness of fit and selected parameter estimates. (**A**) *cadBA* transcription rate ν_m, (**B**) *cadBA* translation rate ν_p, (**C**) maximal turnover velocity of external lysine into external cadaverine v_{max}, (**D**) effective Michaelis constant for lysine turnover, (**E**) *cadBA* mRNA half-life τ_m and (**F**) ratio of maximal number of CadC dimers, C_0 and binding threshold of CadC dimer to the promoter P_{cad}, K_C. The points correspond to local optima in the parameter space, for which the difference between the quantitative model and the experimental data is minimized (see Papers I and II for all details).

The most commonly used regression schemes that minimizes the deviation between model and experiment are maximum-likelihood estimation and least-squares minimization [81]. Under the assumption of uncorrelated, normally distributed measurement errors the likelihood L of parameter set θ is given by the probability to observe the data set y in (2.1), consisting of T time points measured in M observables:

$$L(y|\theta) = \prod_{i=1}^{T}\prod_{j=1}^{M} \frac{1}{\sigma_{ij}\sqrt{2\pi}} \exp\left(-\frac{(y_{ij} - g_i(\mathbf{x}(t_j,\theta),\theta)^2}{2\sigma_{ij}^2}\right) . \tag{2.4}$$

Incidentally, under the mentioned conditions maximizing of L is equivalent to minimizing the sum of squared residuals [147], which then follows a χ^2 distribution

$$\chi^2(\theta) = \sum_{i=1}^{T}\sum_{j=1}^{M} \frac{(y_{ij} - g_i(\mathbf{x}(t_j,\theta),\theta))^2}{2\sigma_{ij}^2} . \tag{2.5}$$

The optimal set of parameters θ^* can then be estimated by numerically determining

$$\theta^* = \arg\min\left[\chi^2(\theta)\right] . \tag{2.6}$$

While in the case of a linear model there exist closed form solutions to this minimization problem [64], the more common problem of nonlinear models is frequently ill-conditioned and multimodal, such that gradient-based local optimization methods fail and deterministic or stochastic global optimization methods need to be applied [124].

In the work described here, we chose the most simple approach to the global optimization problem, the so-called multistart strategy. It is based on the idea of using a local optimization method repeatedly with randomly sampled initial parameter sets from a physiologically

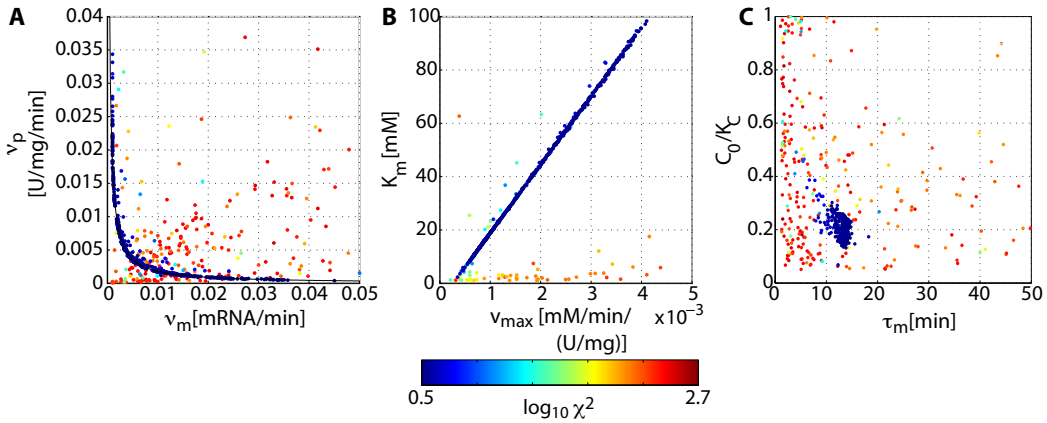


Figure 2.4.: Pairwise correlations between selected fit parameters. For description of parameters please refer to the caption of Fig. 2.3.

relevant subset of the parameter space. In contrast to other global optimization strategies, the multistart method is computationally costly as it will converge to similar local optima several times. However, it has the advantage that it is easy to implement and that the method intrinsically samples the χ^2 "landscape" in the vicinity of the global optimum, providing a means to derive standard errors on the parameter estimates (see below). Fig. 2.3 shows the correlation between the optimized χ^2 values at the local optima and some selected fit parameters of our model for the Cad system (see Paper I for all details). We find for decreasing χ^2 values that some parameters are not well constrained (Fig. 2.3A,B), others are weakly constrained (Fig. 2.3C,D) and some critical parameters are well-constrained (Fig. 2.3E,F). Generally, 'sloppy' parameter sensitivities as in (Fig. 2.3A,B) are often related to the structural or practical non-identifiability of the respective parameters [147], arising from a redundant parametrization of the model or from insufficient amount and quality of the data, respectively. For instance, in the case of the transcription and translation rates ν_m and ν_p of *cadBA* (Figs. 2.3A,B), our experimental data only constrains the total amount of protein, with is determined by the product of both parameters. Consequently, we find that only their product is well-constrained, as reflected in the pronounced anti-correlation of fit parameters with low χ^2 (Fig. 2.4A; *blue spots*). However, it was not possible to subsume transcription and translation into an effective reaction with a single rate constant, since the data demanded a relatively slow timescale (∼12 min) of mRNA degradation which introduced a sharply defined delay in the dynamics (Fig. 2.3E). Another example of sloppy parameter sensitivities involve the parameters for lysine turnover, v_{max} and K_m, which display a linearly correlated variation, indicating that only their ratio is well constrained by our data (Fig. 2.4B).

Derivation of Confidence Intervals

As a measure for the reliability of the parameter estimates, it is necessary to derive a confidence interval $[\theta_i^* - \sigma_i^-, \theta_i^* + \sigma_i^+]$ on each of the best fit values θ_i^*. Unfortunately, due to the complexity of nonlinear optimization problems, there is often no straightforward way to do so. Approximate results can be derived from the curvature of the χ^2 function at its minimum [143], but those are only valid in the limit of infinitely many data and complete parameter

identifiability [62]. Alternatively, likelihood-based confidence intervals can be derived from the borders of the confidence region defined through

$$\{\theta|\chi^2(\theta) - \chi^2(\theta^*) < \chi^2(\alpha, df)\}\,, \tag{2.7}$$

where $\chi^2(\alpha, df)$ is the α quantile of the χ^2 distribution with df degrees of freedom [147]. While these definitions all rely on the normality of measurement errors and are strictly valid only for large data sets, in practice these conditions are often heavily violated. In contrast, bootstrapping methods do not require the prior knowledge about the statistic of interest, that is, the fitted parameters [46]. The basic idea behind bootstrapping is to use the measured data as the basis for resampling a large number of synthetic data sets with the same statistical properties and to run the optimization procedure repeatedly. From the resulting empirical distribution of the parameter estimates one can derive confidence intervals and other statistics [46]. However, also bootstrapping has its limitations, especially if the fitted data do not have equal variance, if the assumed model is not correct or if the data sample is small – facts which can be difficult to evaluate beforehand. Therefore, in the present studies we took a different approach and derived parameter uncertainties directly from the χ^2 "landscape" in the vicinity of the global optimum, as first described by Wall et al. [188]. As mentioned above, the multistart method samples many local optima in the vicinity of the global optimum and from their distribution one can derive the weighed, asymmetric errors σ_+ and σ_- with respect to the optimal parameter values θ^* from the following equations:

$$\sigma^2_{k,+} = \frac{\sum_{i:\theta_{k,i}>\theta_k^{opt}} (\theta_{k,i} - \theta_k^{opt})^2 e^{-\chi_i^2/2}}{\sum_{i:\theta_{k,i}>\theta_k^{opt}} e^{-\chi_i^2/2}}\,, \text{ and } \sigma^2_{k,-} = \frac{\sum_{i:\theta_{k,i}<\theta_k^{opt}} (\theta_{k,i} - \theta_k^{opt})^2 e^{-\chi_i^2/2}}{\sum_{i:\theta_{k,i}<\theta_k^{opt}} e^{-\chi_i^2/2}}\,, \tag{2.8}$$

where $\theta_{k,i}$ is the value of parameter θ_k in the i_{th} fit, θ_k^{opt} is the value of θ_k in the fit with the lowest value of χ^2, and χ_i^2 is the value of χ^2 for the i_{th} fit.

2.4. Paper I: Induction Kinetics of a Conditional pH Stress Response System in Escherichia coli

In the paper 'Induction kinetics of a conditional pH stress response system in *Escherichia coli*', J. Mol. Biol. **393**, 272-286 (2009), by Georg Fritz, Christiane Koller, Korinna Burdack, Larissa Tetsch, Ina Haneburger, Kirsten Jung, and Ulrich Gerland we scrutinized the hypothesis of negative feedback regulation in the Cad system via the end-product cadaverine by combining traditional molecular biology experiments with quantitative mathematical modeling. To that end, we introduced a coarse-grained model for the dynamics of important system variables, namely the *cadBA* mRNA, the CadA activity and the extracellular cadaverine level. From our analysis we found, that the transient response of the Cad system is accurately captured by the model dynamics. Moreover, our results indicate that the precise timing of the transcriptional shut-off is indeed causally related to the increase of external cadaverine above its deactivation-threshold. We further used our quantitative model to predict the dynamics of the system under altered experimental conditions, as, e.g., for a mutant deficient in lysine-sensing. The successful experimental validation of these model predictions corroborates that

the transcriptional shut-off in the lysine-decarboxylase system is neither related to a decrease in the external lysine level nor in the increase of the external pH level. Instead, our analysis puts the feedback inhibition mechanism by cadaverine on firm grounds.

Interestingly, by using parameter estimation techniques described above, our top-down approach allowed us to extract hidden information about relevant biochemical parameters *in vivo*. In particular, the transient response critically hinges on the molecular details of signal integration at the sensor CadC, since the apparent binding affinity and the Hill exponent for the interaction between cadaverine and CadC directly relate the dynamics of external cadaverine to a transcriptional response of the system. Strikingly, the estimated value for the *in vivo* binding constant agreed surprisingly well with previous *in vitro* studies [175]. Moreover, the model predicted an effective Hill exponent of about 3 suggesting a highly cooperative deactivation of CadC by cadaverine.

2.5. Paper II: The Feedback-Inhibitor Cadaverine Suppresses pH Response by Binding to the pH Susceptible Site of CadC in Escherichia coli

In the submitted manuscript 'The feedback-inhibitor cadaverine suppresses pH response by binding to the pH susceptible site of CadC in *Escherichia coli*', by Ina Haneburger*, Georg Fritz*, Nicole Jurkschat, Larissa Tetsch, Andreas Eichinger, Arne Skerra, Ulrich Gerland, and Kirsten Jung (*equal contribution) we set out to investigate the cadaverine-dependent contribution to signal perception in CadC. In the meantime, the crystal structure of the periplasmatic sensory domain of CadC was resolved [47], revealing that CadC is active as a dimer and suggesting a cadaverine binding site within a central cavity of each monomer. Also, it was shown that pH sensing is mediated by a patch of amino acids located right at the dimerization interface [47] and it was speculated that most probably its protonation enables conformational changes that lead to activation of *cadBA* transcription [71]. In our study extensive analysis of several CadC mutants with single, double and triple amino acid substitutions revealed that the predicted binding site only partially accounts for the cadaverine sensitivity of CadC *in vivo*. Instead, a second, pivotal binding site was identified, which is localized directly at the homo-dimerization interface within the pH-responsive patch of amino acids, suggesting that cadaverine binding might interfere with pH detection.

To scrutinize the effect of these potential cadaverine binding sites on the *in vivo* dynamics, we first refined our previously published model by incorporating that CadC is active as a dimer [47] and then analyzed the dynamics of wildtype CadC and two mutants carrying amino acid substitutions in either one of the binding sites. Most strikingly, with our refined model we extracted an effective Hill exponent $n_c = 3.5 \pm 0.6$ for the interaction between cadaverine and CadC in the wildtype. As the Hill exponent gives a lower limit on the number of interacting sites in positively cooperating systems [1], this points to at least four cadaverine binding sites within a CadC dimer. In contrast, both mutants displayed largely reduced Hill exponents of $n_c = 0.5 \pm 0.1$ (mutation in the central cavity) and $n_c = 0.3 \pm 0.1$ (mutation at the dimer interface). Together with the very low cadaverine affinity of the latter mutant, we find that the binding site at the dimer interface is of pivotal importance for cadaverine-sensing *in vivo*, whereas the binding site in the central cavity has less stringent effects of the CadC activity.

Together, these results suggested that the four sites do not modulate CadC dimer activity in a simple, additive manner. Such non-additive effects may result, e.g., from allosteric effects,

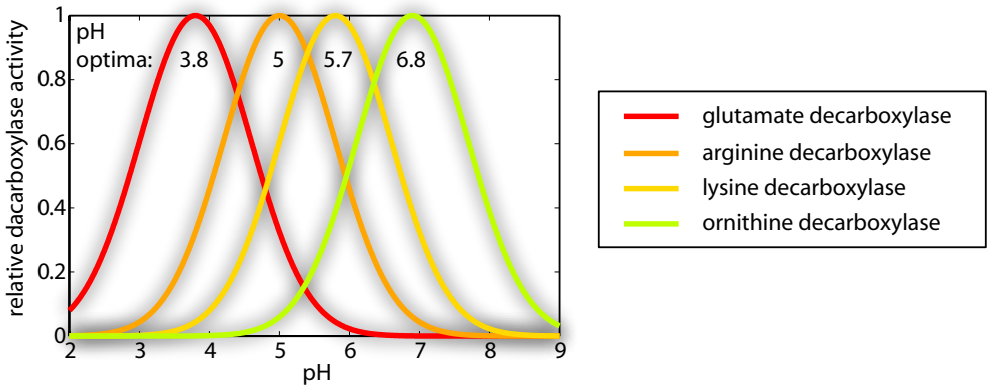


Figure 2.5.: Enzyme activities of various amino acid decarboxylase systems (schematic). Individual enzyme activities are approximated by Gaussians with pH optima and standard deviations drawn to scale [18, 52, 160, 163].

for which binding to a single site induces small conformational changes in CadC wild type, such that binding to the second site is facilitated. Based on our results and all prior knowledge on CadC, we proposed a sequential model for its inactivation by cadaverine: In its activated conformation (low pH and high lysine) CadC forms a homodimer, which grants cadaverine access to the central cavities, but buries the binding sites at the dimerization interface. Upon binding of cadaverine to the central cavities, small conformational changes in CadC dimer expose the binding sites at the dimerization interface. Cadaverine binding to the site at the dimerization interface possibly acts like a wedge that leads to disruption or inhibition of the active CadC dimer conformation, and consequently inactivates CadC.

2.6. Conclusion and Outlook

The tight control and proper orchestration of bacterial stress response modules are of vital importance for bacterial growth. Within this chapter, a comprehensive, systems level characterization of one such module was presented, the results of which shed light on the molecular details of negative feedback regulation via cadaverine the Cad system in *E. coli*. The mechanism of cadaverine perception by CadC augments our understanding of signal integration by one-component systems in general and gives a glimpse at the information-processing capacity of simple and compact polypeptides. However, the physiological need for this negative feedback remains elusive so far. One might speculate that it serves as a "produce-to-demand" mechanism, which takes the external cadaverine level as a proxy for the abundance of CadA and CadB, such that the negative feedback by cadaverine asserts homeostatic control of their levels. In that, negative feedback regulation via cadaverine could minimize the cost of *cadBA* induction and might save cellular resources for the investment in other cell-protecting processes. In the future it will be interesting to test this hypothesis, for instance, by performing competition experiments between wildtype cells and mutants impaired in cadaverine sensing.

More broadly, analogous quantitative studies of the other amino acid decarboxylase systems (glutamate, arginine, ornithine, etc.) could reveal important insights into how these modules are orchestrated in the complex environment of their host. From the non-overlapping pH-optima of the individual decarboxylase activities (Fig. 2.5), it is clearly visible that this battery of acid stress response modules has the potential to protect against a wide spectrum

of internal pH values. Therefore, it seems plausible that in the natural environment of *E. coli* (where typically multiple amino acids are present) these systems are often induced in parallel. It will then be interesting to probe whether there exists a temporal hierarchy in the induction of these modules upon gradually decreasing external pH and to test, whether cells use some sort of cross-activation between individual modules to allow for an adaptive prediction of decreasing internal pH values.

3. Memory in Microbes: On Schemes of Sequential Gene Regulatory Logic

Through all kingdoms of life transcriptional networks equip cells with a remarkable behavioral repertoire. Not only do they allow for appropriate responses to present cellular and environmental signals, but, to a certain extent, also enable microbes and cells in multicellular arrangements to "memorize" the history of these cues in form of distinct gene expression states. Examples include the decision between lysis and lysogeny at the time bacteriophage lambda infection [166], the multistability in various sugar utilization systems of *Escherichia coli* [119, 130, 134, 161] and *Saccharomyces cerevisiae* [2], as well as the differentiation of stem cells into specialized tissues during eukaryotic development [26, 37, 42, 65, 138, 181, 190]. Clearly, such history-dependent behavior can provide selective advantages to an organism and thereby favor the evolution of memory in natural genetic circuits. However, also in forward engineering approaches, in which synthetic gene circuits are designed for designated functions, the ability to collect and store information at specific points in time and space can be highly desirable [5, 61, 79, 91].

In analogy to electronic circuits, genetic networks with memory are often referred to as sequential logic circuits [57, 70, 114, 118, 196]. Unlike combinatorial circuits, sequential circuits hold a stable internal state and, as illustrated in Fig. 3.1, generate a history-dependent output: When external signals are applied to the sequential circuit, a combinatorial circuit integrates the primary (external) inputs and the current state of the memory element. In addition to the primary output also a secondary output is generated, which is fed back into the memory device as a next internal state. In that, sequential circuits hold information about the history of their input signals. In the same way that logic gates are the building blocks of combinatorial circuits, latches and flip-flops are the building blocks of sequential circuits. Previously, the general schemes of combinatorial transcription logic have been studied both theoretically [16, 17, 23] and experimentally [35, 68, 83, 97]. Although a number of recent case-studies focussed on transcriptional circuits with memory [2, 5, 31, 57, 61, 70, 79, 114, 194], a global view on the design principles of sequential transcription logic is still missing. In close collaboration with Patrick Hillenbrand (LMU Munich) such a comprehensive view was developed, and it is the focus of this chapter to give a simple and transparent classification of the latches and flip-flops for sequential logic design.

First, the concepts of purely combinatorial logic are introduced and some of the genetic implementations of gene regulatory logic gates are briefly reviewed (Section 3.1). Then I focus on sequential logic in gene regulation by discussing some of the "memory devices" which can be leveraged for biological information storage and by giving a general classification of sequential logic elements (Section 3.2). In particular, it will be analyzed how the molecular repertoire of bacterial gene regulation can be leveraged to find compact and robust genetic implementations of these basic storage elements, which are functional in the presence of the delays inherent to gene expression (Section 3.3). Here it turns out that the combination of protein-protein interactions and cis-regulatory protein-DNA interactions are sufficient in-

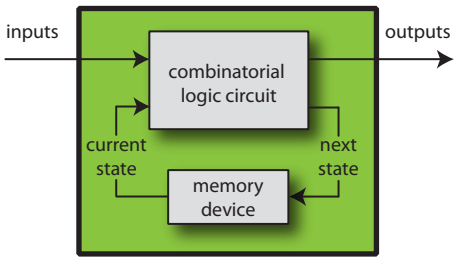


Figure 3.1.: Scheme of a sequential logic circuit. In contrast to pure combinatorial logic, the output of a sequential circuit depends both on the present input signals as well as on the current state of an internal memory device. In addition to the external output also a secondary output signal is generated, which, after a component-specific delay, is fed back into the memory device as next internal state.

gredients to implement complex sequential logic, without the need for further regulatory mechanisms. Finally, I discuss the significance of sequential logic in natural circuits (Section 3.4) and conclude with some potential applications to synthetic biology (Section 3.5).

3.1. Combinatorial Transcription Logic

In combinatorial logic circuits all outputs are directly related to the current combination of values on the input signals, i.e., they are memoryless. The most elementary combinatorial logic operations are implemented by so-called "logic gates", which relate two input signals to a single output value. For instance, an AND gate generates a positive output (corresponding to a logic 1) only if both input signals are applied at the same time and yields 0 else. Generally, such input-output relations are represented in form of a truth table, as shown in the table at top of the figure. As there are four rows within the truth table, i.e., four distinct combinations of the input signals X and Y, there exists a maximal number of $2^4 = 16$ input-output relations. Leaving aside trivial gates in which the output only depends only on one (or even on none) of the inputs, and considering the symmetry between the input signals X and Y, only 8 gates with distinct input-output relations remain (see Fig. 3.2; *top panel*). However, from digital circuit theory it is known that any boolean function with arbitrary number of inputs can be further reduced to one of two *canonical forms*, which only use cascaded combinations of AND, OR and NOT gates to perform the same logic operation [84].

While electronic circuits indeed feature such "deep" architectures with many layers of simple but fast nodes, gene-regulatory networks typically cannot afford long cascades of genes regulating the activity of other genes. Instead, the slowness of protein expression and the limited number of genes makes it more reasonable adopt a "broad" architecture integrating complex computations such as the "exclusive OR" (XOR) or the "equivalence" (EQ) gate into a single node. On a molecular level, specific protein-DNA interactions and weak glue-like interactions between regulatory proteins proved to be sufficient ingredients to implement regulatory logic functions of increasing complexity [23]: By appropriately selecting the strengths and arranging the relative positions of the relevant protein-binding DNA sequences in the cis-regulatory region, each of the 8 elementary logic operations can be carried out by a single gene [23]. For instance, a logic AND function can be constructed by choosing weak binding sites for both X and Y and placing them adjacent to each other (see figure) such that each transcription factor alone cannot bind to its site, but when both are present binding occurs

input X	input Y	AND	NAND	OR	NOR	XOR	EQ	`Lac'	`NOT Lac'
0	0	0	1	0	1	0	1	0	1
0	1	0	1	1	0	1	0	1	0
1	0	0	1	1	0	1	0	0	1
1	1	1	0	1	0	0	1	0	1

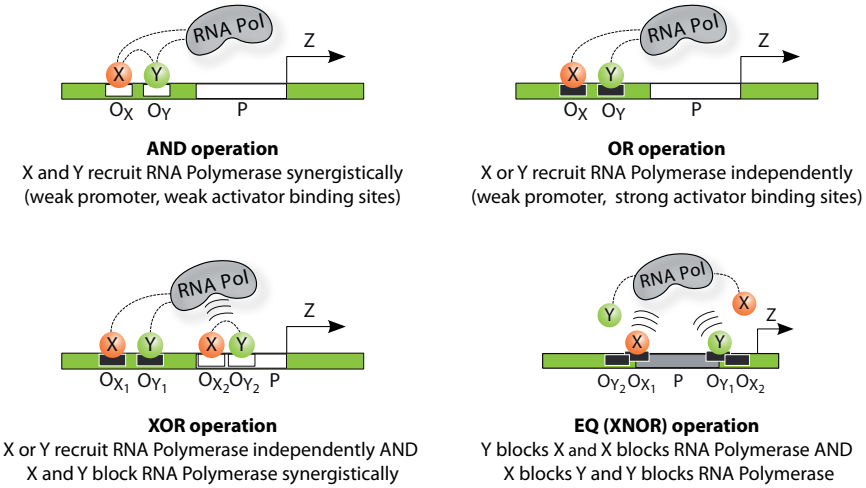


Figure 3.2.: Truth tables (*top panel*) and exemplary *cis*-regulatory implementations of some combinatorial logic gates.

with the help of an additional cooperative interaction. Once X and Y are bound, they recruit RNA Polymerase (RNA Pol) to the weak promoter site and activate transcription.

While this scheme is useful for some of the elementary gates, there are limitations to the control functions one can implement by using only the two ingredients of regulated recruitment, as illustrated with the EQ gate (see Fig. 3.2). A strong promoter is required here to turn the gene ON when neither of the TFs are activated, whereas repression is needed under multiple conditions (i.e., when X is activated and Y is not, and vice versa). Given the small size of the promoter region, it is difficult to implement both repressive conditions by the direct physical exclusion of RNAP. An effective strategy to overcome promoter overcrowding is repression from a distance, e.g., through DNA looping. In fact, it has been shown that the distal repression and activation schemes can be regarded as natural representations of the *conjunctive* or *disjunctive normal forms* of logic, respectively, and are therefore suited to code for any given logic function [23].

3.2. Sequential Logic in Gene Regulation

3.2.1. Bits of Biological Information Storage

Both prokaryotic and eukaryotic cells store information on various timescales. For short signal durations covalent modifications of preexisting proteins are frequently used as rapidly addressable and yet stable "bits" of biological signaling cascades. In phosphorylation cascades for instance, the activated (phosphorylated) form of a protein often acts as a kinase for the

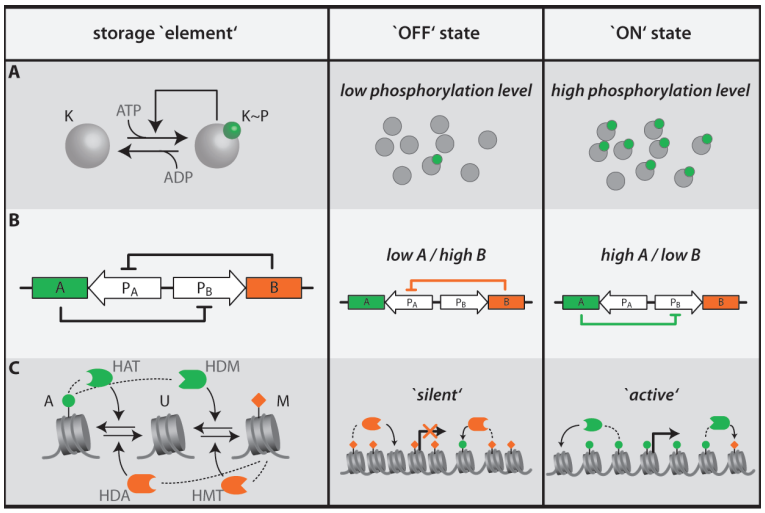


Figure 3.3.: Basic storage elements in gene regulatory circuits. (**A**) Positive feedback in a single enzyme system with autokinase activity. (**B**) Genetic toggle switch of two mutually repressing genes. (**C**) Bistable nucleosome modification (for details see text).

phosphorylation of an earlier protein in the cascade, effectively closing a positive feedback loop. Even for the simple example of a single enzyme with autokinase activity (Fig. 3.3A) this positive feedback can lead to bistable behavior [110]. But also in more complex cascades, such as in the p42 mitogen-activated protein kinase (MAPK) cascade involved in the maturation decision of *Xenopus* oocytes, positive feedback loops produce an actively maintained memory of a transient stimulus and provide an explanation for the irreversibility of maturation [195]. Similar mechanisms are also found in neurons, in which bistable phosphorylation signals set the basis for long-term potentiation and depression of synapses, and are thus pivotal for synaptic plasticity [132].

Interestingly, in neurons a later phase of long-term depression requires the synthesis of proteins [4, 109], and it has been speculated that long-term memory might be preserved by bistable expression kinetics of those genes [132]. In fact, the expression levels of bistable genetic switches can be inherited to daughter cells, allowing cells to store information for several cell generations [2, 130, 134, 161]. A classical example is the genetic toggle switch [61], which is based on two mutually repressing genes A and B (Fig. 3.3B). If the expression of gene A is high initially, it represses the expression of gene B and the system stably remains in the (A high, B low state) and *vice versa*. In that, it functions very similar to a toggle switch in electronic circuits. However, while the electronic toggle switch maintains its state as long as desired, the stability of the genetic toggle switch is typically limited by stochastic gene expression, which can lead to noise-induced "memory-loss" [2].

Establishment of cellular memory and its faithful propagation for many cell generations is pivotal for successful development of multicellular organisms. As pluripotent cells differentiate, choices in cell fate need to be inherited and maintained by their progeny throughout the lifetime of the organism. A major factor in this process is the remodeling of chromatin structure: DNA methylation, histone modifications and many other types of alterations of the genetic code allow eukaryotic cells to turn genes on and off as needed. Interestingly, positive feedback can arise if nucleosomes that carry a particular modification recruit (directly or indirectly) enzymes that catalyze similar modification of neighboring nucleosomes

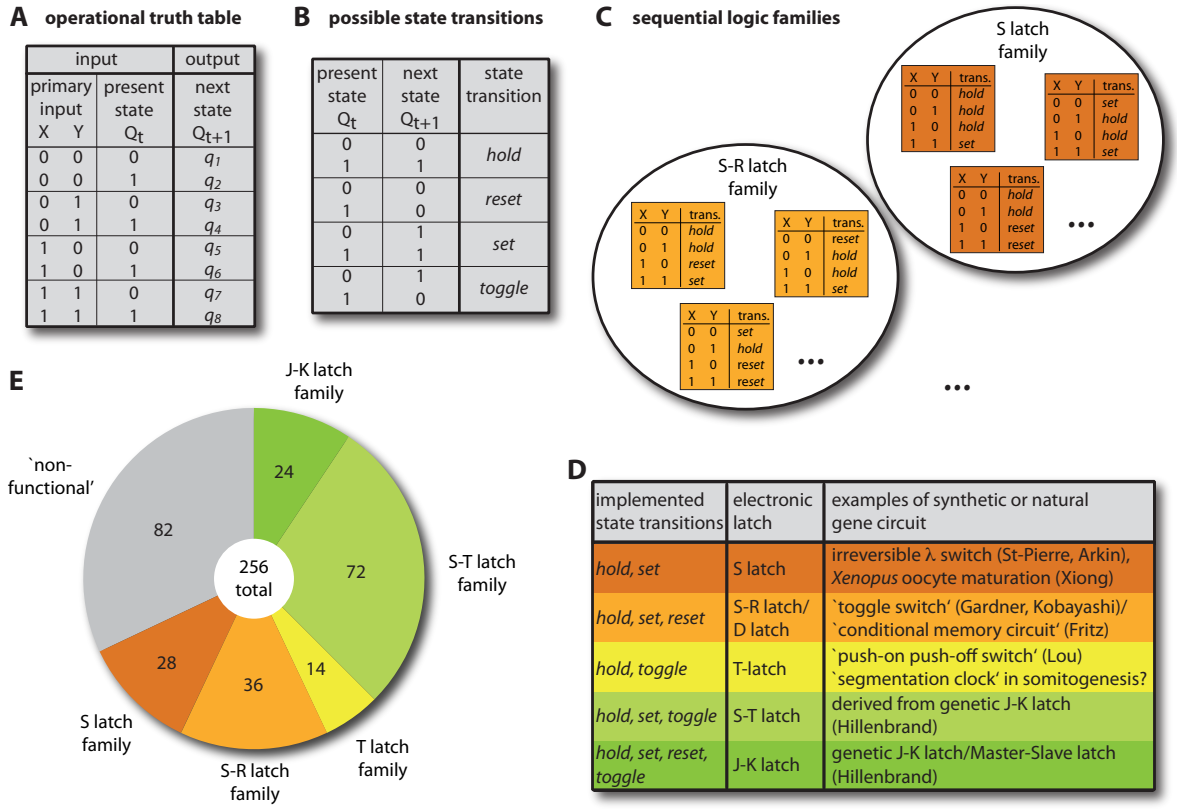


Figure 3.4.: Classification of 1-bit sequential logic elements into functional families. (**A**) The boolean function of a 1-bit sequential logic circuit is described by an operational truth table. Here the output, i.e., the next state Q_{t+1} of the device, depends on both the primary inputs X and Y as well as on its current state Q_t. (**B**) Four possible transitions of the internal state. (**C**) Sequential logic elements are classified into functional families according to their repertoire of implemented state transitions. (**D**) The 5 families of functional sequential logic devices. (**E**) Individual frequencies of the sequential logic families among all possible 1-bit memory elements.

[43, 144]. As illustrated in the scheme of Fig. 3.3C (adapted from ref. [43]), nucleosomes can exist in unmodified (U), methylated (M), and acetylated (A) states. These forms are actively interconverted by modifying and demodifying enzymes that are recruited by the modified nucleosomes themselves. For instance, methylated nucleosomes recruit histone deacetylases (HDAs) and histone methyltransferases (HMTs) (*red symbols* in Fig. 3.3C), which act on surrounding nucleosomes and thereby allow a cluster of nucleosomes to maintain itself stably in a methylated, transcriptionally inactive (silent), state [43]. Similarly, the recruitment of histone demethylases (HDMs) and histone acetyltransferases (HATs) (*green symbols* in Fig. 3.3C) by acetylated nucleosomes promotes stable, transcriptionally active, clusters of acetylated nucleosomes [43].

3.2.2. Classification of Sequential Logic Elements

For a systematic analysis of sequential logic, I now focus on the most elementary "latches", which constitute the building blocks for more complex systems with memory. Generally, a

latch holds only a single bit of information, that is, its internal memory can either be 'OFF' or 'ON', corresponding to a logic 0 or 1, respectively. In its simplest form, the output of the latch equals the state of its internal memory device and its boolean logic function can be represented in an operational truth table (Fig. 3.4A) – in analogy to the truth table of a combinatorial circuit (Fig. 3.2; *top panel*). Here both the external input signals X and Y, as well as the current state Q_t of the memory device determine the next state Q_{t+1} of the sequential logic element. The combination of 3 binary input signals leads to $2^3 = 8$ possible input conditions, i.e., to 8 rows in the operational truth table, each of which is associated with a binary output. Hence, there exist as many as $2^8 = 256$ distinct operational truth tables, which set the basis for the most elementary class of 1-bit sequential logic elements. However, from combinatorial logic it is known that the number of functionally distinct gates is reduced significantly by taking into account the symmetry between input signals and by leaving aside gates with trivial output (see above). In the following, I show that similar considerations narrow down the number of 1-bit sequential logic elements as well.

To this end it is important to note that an arbitrary combination of external inputs to a 1-bit memory element can only cause four distinct transitions of the internal state (Fig. 3.4B): The device either keeps its current state (*hold*), is set into the 'ON' state (*set*), is reset to the 'OFF' state (*reset*), or it changes its state into the complement of the current state (*toggle*). With this definition, the operational truth table in Fig. 3.4A can be rewritten in form of a conventional truth table with only X and Y as inputs, but with the corresponding state transition as output (cf. Fig. 3.4C). This condensed form of the operational truth table contains the internal state of the memory implicitly, and puts the focus on the various ways in which the external input signals manipulate the state of the device. Based on this operational view we classify all 1-bit sequential logic elements into different families, depending on their repertoire of implemented state transitions (Fig. 3.4C,D). For instance, the S-R latch family contains all latches that only implement the *hold*, *set* and *reset* transitions (Fig. 3.4C). Since the four input combinations of X and Y code only for three state transitions in this case, two input combinations have to code for the same transition. This redundancy in the number of implemented transitions, together with the various permutations in which the four input combinations encode the transitions, leads to 36 members within the S-R latch family (Fig. 3.4E).

Common to all functional families is, that at least one row within the operational truth table codes for the *hold* command. In fact, circuits unable to hold their internal state are non-functional in the sense that they produce a purely combinatorial (memoryless) output and continuously alter their state. Likewise, circuits which exclusively implement the *hold* command are not of practical use, since their state cannot be addressed. Based on these considerations 82 out of the 256 latches are classified as non-functional (Fig. 3.4E). Moreover, since the state transitions *set* and *reset* are symmetrical in the sense that they can be substituted by either defining the internal state or its complement as the 'ON' state, we combine latches that contain only one of the two transitions into a single family. For instance, the S latch family contains latches that implement either the *hold* and *set* or the *hold* and *reset* transitions (Fig. 3.4C). In summary, we distinguish between 5 functional families of 1-bit sequential logic elements, which are listed in Fig. 3.4D together with their repertoire of implemented state transitions.

3.3. Molecular Implementation of Basic Genetic Latches

The ability to store information lies at the heart of every sequential logic circuit. In digital electronics, most one bit latches are constructed from two cascaded NOT gates with feedback, effectively implementing a positive feedback loop of each gate on itself. From nonlinear dynamical systems it is well known that positive (or double-negative) autoregulation can lead to bistable behavior also in biological circuits [51], and both prokaryotes and eukaryotes make active use of such epigenetic forms of information storage (see Fig. 3.3 and reference [24] for a recent review). To illustrate how members of the different sequential logic families can be implemented from these basic memory bits, we focus our analysis on well studied mechanisms of prokaryotic gene regulation. For instance, a classical example of a bistable circuit is the genetic toggle switch [61], which is based on two genes A and B mutually repressing their activities through the concentrations of their gene products, the transcription factors (TFs). If the expression of A is high initially, it represses the expression of B and the system stably remains in the high A/low B state and *vice versa*.

In order to address the state of the genetic toggle switch, an interface to external control signals as well as to the current state of the device is essential (cf. Fig. 3.1). In digital electronics, combinations of AND and NOT gates are typically sufficient to provide the combinatorial control required for the implementation of all possible 1-bit latches [84]. Correspondingly, the cascaded expression of several auxiliary TFs might serve as "regulatory front end" to the genetic toggle switch, by integrating all input signals in their *cis*-regulatory regions and by sensitizing the toggle switch for their output. Alternatively, signal integration can be performed already at the protein level, e.g., through homo- and heterodimerization between different TFs, which achieves the desired characteristics without any intermediate genes, thereby avoiding many potential problems associated with their expressions (e.g., time delay and stochasticity). For instance, heterodimerization between TFs X and Y and subsequent repression of a target gene by XY provides the functionality of a molecular NAND gate, as has been demonstrated by engineered variants of the LacI/O_{lac} pairs [198]. Likewise, recent examples in *Escherichia coli* show that prokaryotic cells also make active use of heterodimeric transcription control *in vivo* [30, 184].

3.3.1. Basic Model

Extending on our previously published model for a conditional memory circuit [57], we consider the scheme in Fig. 3.5A as a template for our design of 1-bit gene regulatory latches. Here the genetic toggle switch (*upper part*) serves as memory device for all latches, whereas the interface to, and the interactions within, the regulatory front end (*lower part*) assign the logic in which the toggle switch responds to external input signals and thus specifies the behavior of an individual latch. More specifically, we take the expression rates α_X and α_Y of genes X and Y as input signals, and allow homo- and heterodimerization among their gene products X and Y, as well as with the proteins A and B of the toggle switch. In addition to the repressor binding sites for the dimers A_2 and B_2 in the promoter regions of the toggle switch (two binding sites for each species are required to achieve bistability [57]), two auxiliary repressor sites O_{R_A} and O_{R_B} are introduced downstream of the transcription start sites of genes A and B, respectively. Supposing that the DNA-binding domains of all TFs recognize disjunctive target sequences on the DNA (see e.g. references [77, 198]), the choice of the sequences of O_{R_A} and O_{R_B} makes the toggle switch sensitive only for a specific pair of

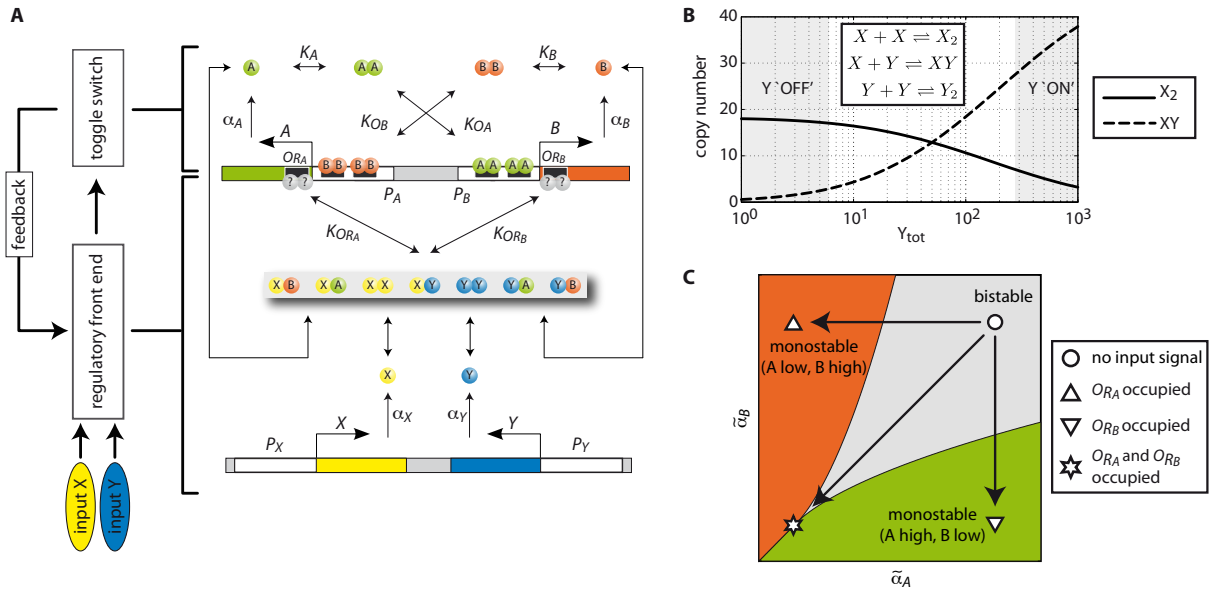


Figure 3.5.: (**A**) Template scheme for the implementation of 1-bit gene regulatory latches. The K_i denote equilibrium dissociation constants, P_i the promoter of gene i and α_i the corresponding protein expression rate. For simplicity, synthesis and degradation of mRNA species of all TFs as well as degradation of TF molecules are not depicted. (**B**) Shift of dimerization equilibrium upon changes in the total number of proteins Y_{tot} for the reactions depicted in the inset (for clarity, the abundance of Y_2 is not shown). For all reactions the equilibrium dissociation constant is $K = 10\,\mathrm{nM}$ and the total number of proteins X is fixed ($X_{tot} = 50$). (**C**) State diagram of the genetic toggle switch, showing its stability as a function of the effective expression rates $\tilde{\alpha}_A$ and $\tilde{\alpha}_B$ of genes A and B, respectively.

homo- or heterodimers R_A and R_B in the regulatory front end. In addition, the affinities of the TF monomers for each other determine the protein concentrations at which the individual homo- and heterodimers form. Hence, changes in TF concentration can shift the equilibrium from one homo- or heterodimer-species to the other, permitting a concentration-dependent control over the state of the toggle switch, as will be shown now.

The behavior of the toggle switch in response to the input signals X and Y can be best understood within a reduced deterministic description which only considers the total concentrations of proteins A and B as dynamic variables. Such a description assumes that all biochemical processes which do not change the total concentrations A_{tot} and B_{tot} are sufficiently rapid to remain equilibrated at almost all times. The net change of A_{tot} and B_{tot} due to protein synthesis and degradation then follows rate equations of the form

$$\frac{dA_{tot}}{dt} = \alpha_A P_A(B_2, R_A) - \lambda A_{tot} \tag{3.1a}$$

$$\frac{dB_{tot}}{dt} = \alpha_B P_B(A_2, R_B) - \lambda B_{tot}\,. \tag{3.1b}$$

Here, it is assumed that protein degradation occurs at the constant rate λ, while the synthesis of proteins A and B is regulated. Their maximal synthesis rates are denoted by α_A and α_B and the form of the regulatory control is described by the promoter activity functions $P_A(B_2, R_A)$ and $P_B(A_2, R_B)$. They determine the fraction of time the promoter is not blocked by a

repressor and thereby free to bind RNA polymerase [16, 17]. In the simplest case of non-cooperative binding of the dimeric repressors to their operator sites we have

$$P_A(B_2, R_A) = \left(1 + \frac{B_2}{K_{O_B}}\right)^{-2} \left(1 + \frac{R_A}{K_{R_A}}\right)^{-1} \tag{3.2a}$$

and similarly,

$$P_B(A_2, R_B) = \left(1 + \frac{A_2}{K_{O_A}}\right)^{-2} \left(1 + \frac{R_B}{K_{R_B}}\right)^{-1}, \tag{3.2b}$$

where the K's denote the equilibrium dissociation constants for the dimer-operator interaction. Note that the dimer concentrations A_2 and B_2 in Eqs. (3.2) must be expressed in terms of the total protein concentrations A_{tot} and B_{tot} to close the rate equations (3.1). Similarly, the concentrations of the control proteins, R_A and R_B, can be functions of A_{tot} and B_{tot} as well as of X_{tot} and Y_{tot}, which we use to quantify the strengths of the input signals. For instance, in a simple scenario, in which gene A is regulated by $X_2(= R_A)$ and gene B is regulated by $XY(= R_B)$, both control dimers are absent if X_{tot} is low. Only at high X_{tot}, the dimerization equilibrium shifts from mostly forming X_2 at low Y_{tot} to forming mostly XY at high Y_{tot} (Fig. 3.5B).

Consequently, without input signals ($X_{tot} = Y_{tot} = 0$), the second factors on the right hand side of Eqs. (3.2) disappear and the circuit behaves like the regular toggle switch [61]. To assert that the circuit implements the *hold* command in the absence of input signals, the maximal promoter activities α_A and α_B must be similar and both sufficiently strong (Fig. 3.5C; *bistable regime*). In the presence of input signals, the maximal promoter activities effectively are replaced by $\tilde{\alpha}_A = \alpha_A/(1 + R_A/K_{O_A})$ and $\tilde{\alpha}_B = \alpha_B/(1 + R_B/K_{O_B})$. Hence, variation in the concentrations of the control dimers, R_A and R_B, effectively change the maximal promoter activities, and thus can be interpreted as regulated shifts within the state space of the toggle switch. Input signals causing the formation of R_A alone, lead to occupation of O_{R_A} and drive the circuit into the monostable OFF regime (Fig. 3.5C; *red area*). Subsequent removal of the input signal completes the *reset* transition by allowing the toggle switch to relax back to the bistable regime while preserving the OFF state. Likewise, the transient formation of R_B alone implements the *set* transition and turns the state of the circuit ON. However, simultaneous formation of R_A and R_B must be avoided, as it drives the toggle switch into an ill-defined (A low, B low) state (cf. Fig. 3.5C), which, after signal removal, results in the random relaxation into either the ON or the OFF state.

3.3.2. Exploring the Design Space

Next, we scrutinize how the combinatorial repertoire of homo- and heterodimerization in the regulatory front-end can be leveraged to perform the logic operations required by individual members of each sequential logic family. As mentioned above, the genetic sequence of the binding sites O_{R_A} and O_{R_B} make the toggle switch sensitive towards individual homo- and/or heterodimers from the regulatory front-end. In the following, we will gradually increase the complexity of signal integration within the regulatory front-end by allowing different dimer species to bind to the two binding sites. In order to facilitate our discussion, we will rename the proteins X and Y encoding the input signals according to their function within the individual latch.

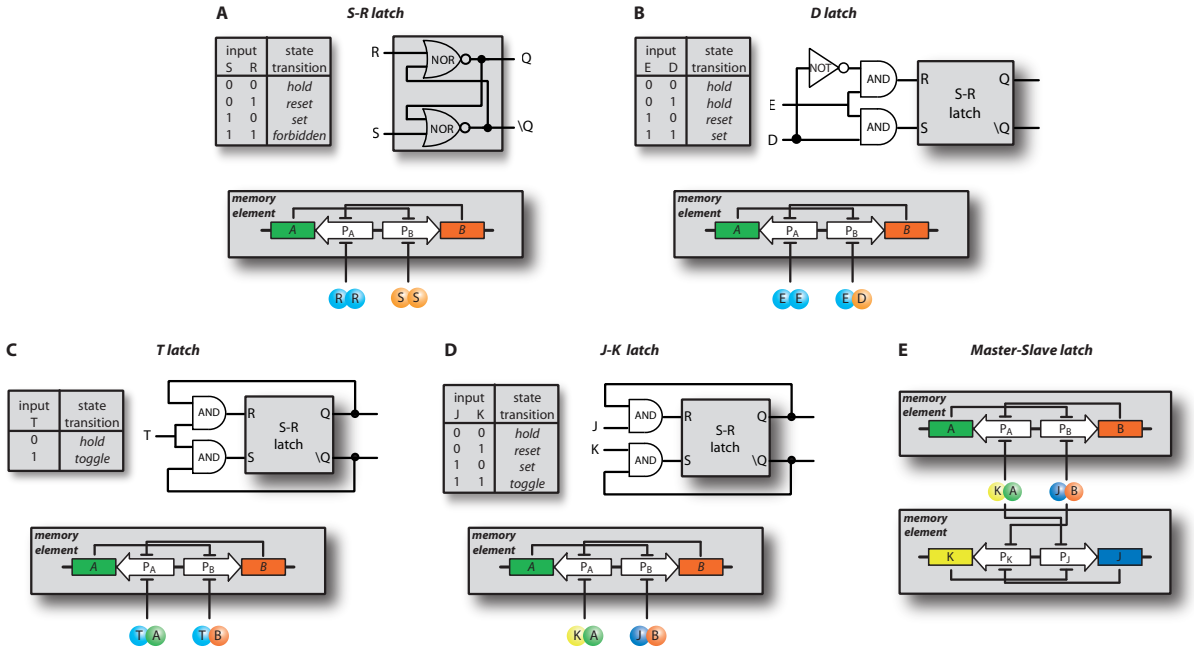


Figure 3.6.: Operational truth tables, most common boolean representation and possible genetic implementation of the S-R latch (**A**), the D latch (**B**), the T latch (**C**) and the J-K latch (**D**). The Master-Slave latch (**E**) has the same operational truth table as the J-K latch.

S-R latch. In its most simple form, only the homodimers R_2 (= X_2) and S_2 (= Y_2) of the input signals specifically bind to the toggle switch and the circuit acts similar to an S-R latch (Fig. 3.6A): Clearly, if only one of the input signals is expressed, the toggle switch is either driven into the A off/B on state by binding of the repressor R_2 to O_{R_A} (*reset*) or driven into the A on/B off state by binding of the repressor S_2 to O_{R_B} (*set*). However, if both input signals are applied at the same time, the the repression of both promoters forces the circuit into an ill-defined A off/B off state, which is expected to relax randomly into either one of the states - depending on stochastic fluctuations at low molecule numbers. Therefore, this combination of input signals is - as in digital electronics - referred to as *forbidden*, as it leads to unpredictable behavior of the circuit (see Fig. 3.6A; *operational truth table*).

To test the behavior of the circuit *in silico*, a full deterministic model was set up according to the reaction scheme in Fig. 3.5 including transcription, translation and degradation of all TF species and their corresponding mRNAs (not depicted), as well as protein (hetero-) dimerization and protein-DNA interaction. The corresponding reaction rates and equilibrium constants were chosen from a physiological range of typical promoters and TFs, allowing us to probe the circuit dynamics for realistic parameters known from bacterial gene regulation (see Appendix B for the full reaction scheme and a comprehensive list of parameter values). However, since a rapid response of the memory can be obtained only when degradation is rapid [61], we assume that all proteins are actively degraded with half-lives of 5 min, which is usually achieved in synthetic gene circuit experiments by SsrA tags [9, 49, 59].

The dynamical response of the proposed genetic S-R latch upon the different state transitions is probed in Fig. 3.7A. Here the top panel displays the dynamically changing concentrations of the homodimers R_2 and S_2 in the regulatory front end, the transcription rates of which are taken as external control parameters. During the first two hours both input signals

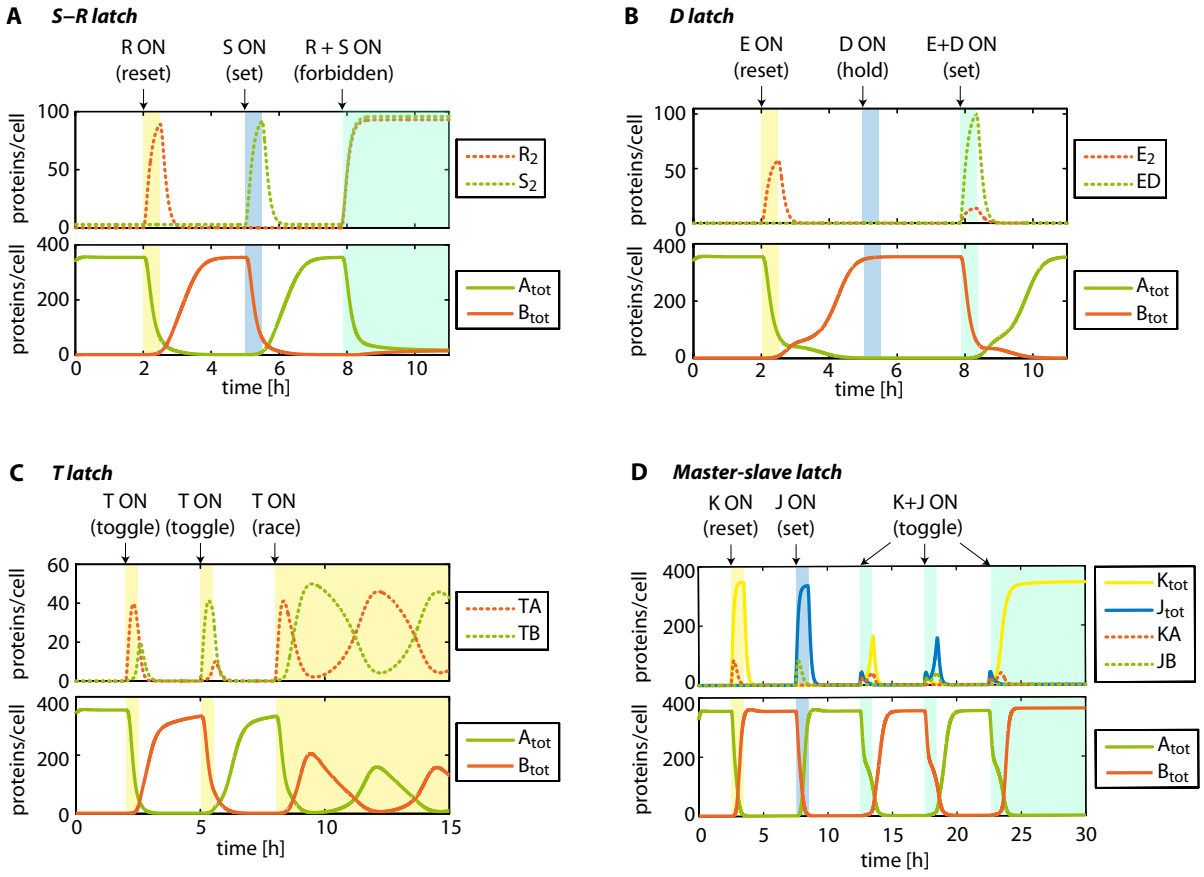


Figure 3.7.: Dynamical test of the state transitions in the S-R latch (**A**), the D latch (**B**), the T latch (**C**) and the Master-Slave latch (**D**).

are OFF and the toggle switch (Fig. 3.7A; *bottom panel*) keeps its initial A ON/B OFF state. As soon as transcription of R is turned ON ($t = 2\,\mathrm{h}$), R_2 dimers are produced and quickly reach the threshold for repression of transcription from the promoter P_A. Consequently, proteins A are degraded at a timescale determined by the SsrA tag and once its concentration falls below the repression threshold for the promoter P_B, the synthesis of B initiates and further represses the transcription of A from P_A. At that time the toggle switch has successfully flipped its state and stably remains in the A OFF/B ON state even after removal of the *reset* signal 30 min after its first appearance (Fig. 3.7A; *bottom panel*). Similarly, the reverse switching from the A OFF to the A ON state is triggered by induction of the S protein (*set* signal) for 30 min. However, as mentioned above, the simultaneous presence of both R and S forces the toggle switch into an ill-defined A OFF/B OFF state (Fig. 3.7A; *cyan area*), which is – in the presence of intrinsic noise – expected to evoke random relaxation into either the A ON or the A OFF state upon signal removal. Therefore, in the design of circuits with S-R latches it is crucial to ensure that this forbidden input does not occur.

D latch. In practice it is more convenient to use a D latch instead of an S-R latch, since it is free of ambiguous input conditions (Fig. 3.6B; *operational truth table*). Here the inputs E and D can be thought of as "enable" and "data" signal, respectively. If the enable signal is

'OFF', the latch holds its state irrespective of the data signal. Conversely, if the enable signal is 'ON', the latch takes the state of D. In digital electronics the D latch can be constructed from an S-R latch by a simple rewiring of the input signals, as illustrated in the digital wiring diagram of Fig. 3.6B. If $E = 0$ both outputs of the AND gates are 0 and thus the S-R latch holds its state, whereas if E = 1 the state of D determines whether the set or the reset signal gets activated.

To overcome the limitation of forbidden input signals in the genetic toggle switch, we previously proposed a genetic implementation of a D latch [57] – the conditional memory circuit (Fig. 3.6B; *genetic wiring diagram*). However, instead of adding additional transcription factor genes performing the AND gate logic required in the D latch, the conditional memory circuit uses heterodimerization between the two input signals as a particularly compact and rapid implementation of regulatory front end: The protein E can form homodimers E_2 and repress the transcription of gene A. It can also bind to D to form heterodimers ED which repress the transcription of gene B. Qualitatively, we expected that if E is absent, neither E_2 nor ED can form: in this state, the existing memory of the circuit is maintained regardless of the state of the level of D. When a significant amount of E is present, it will mostly form homodimers E_2 at low concentrations of D, so that gene A is repressed and the switch is forced into the 'OFF' state. Conversely, when D is highly expressed, mostly the heterodimers ED will form and force the switch into the 'ON' state. This qualitative expectation was verified by simulation of the circuit dynamics under realistic parameters as for the S-R latch (Fig. 3.7B). Here it is clearly visible that the *set* and *reset* transitions are only triggered in the presence of the *enable* signal and that in its absence the toggle switch holds its state irrespective of the signal carried by D. Hence, the heterodimer-logic within the regulatory front end provides a compact and efficient means of conditioning the decision to remember or ignore specific signals upon the presence of hierarchical signals.

T latch. One common requirement in digital circuits is counting, both forward and backward. Although there exist many forms of binary counters, they are fundamentally very similar and rely on the ability to flip the current state of the memory element into its complement [84]. The most basic device with this functionality is the *toggle latch* ("T latch"), which implements both the *hold* and the *toggle* command (Fig. 3.6C). The crux of realizing the *toggle* transition is that a single signal must set the state of the latch either 'ON' or 'OFF', depending on the current state of the device. In that, the *toggle* transition is – together with the much simpler *hold* command – the only genuine sequential logic operation. As illustrated in the digital wiring diagram of Fig. 3.6C, the boolean design is based on an S-R latch as a central memory unit and is complemented with two upstream AND gates, each of which is connected to one input of the S-R latch. The two parallel AND gates integrate both the primary input T of the circuit (the *toggle* signal), as well as the current state Q and its complement $\backslash Q$ of the S-R latch. If the *toggle* signal is OFF ($T = 0$) the circuit keeps its state, since both AND gates provide a logic zero to the S-R latch. If the *toggle* command ($T = 1$) is applied to the circuit, the front end triggers the appropriate *set* or *reset* input of the S-R latch, such that it switches its state into its complement. For instance, if Q was 'ON' previously, the feedback of Q on the front end activates the *reset* signal and the S-R latch switches to the 'OFF' state. Similarly, the *toggle* command switches Q to the 'ON' state if the T latch was 'OFF' before.

The genetic analogue to the digital T latch can be designed in the same spirit as above, by

making the toggle switch sensitive towards heterodimers TA and TB, which form between the TFs of the toggle switch and the regulatory front end (Fig. 3.6C; *genetic wiring diagram*). However, quantitative analysis with our *in silico* model shows that this circuit quickly relaxes into the ill-defined A OFF/B OFF state in the presence of the *toggle* signal – as does the simple S-R latch. The reason behind this monostable behavior lies in the timescales involved in the circuit dynamics: If the circuit is, say, in the A ON state initially, the expression of T leads to the formation of TA and causes the repression of P_A, as desired. As soon as transcription of B initiates, however, TB forms instantly and auto-represses the establishment of the new B ON state. This behavior is unlike the dynamics of an electronic T latch, in which component-specific delays induce stable oscillations between the ON and the OFF state in the presence of a continuously applied toggle signal. In contrast, in our genetic design the compact architecture of the regulatory front end leads to a feedback that is much faster than the dynamics of the toggle switch itself.

To circumvent this rapid feedback, the expression of two auxiliary TFs might serve as a much slower implementation of the regulatory front end, by integrating all input signals in their cis-regulatory regions and by sensitizing the toggle switch for their output. Specifically, the two AND-operations required for the design of the T latch can be carried out by cooperative activation of a "reset gene" R (and "set gene" S) by A and T (and B and T): If the conditions for expression of R are met (T is high and the A is on), it represses expression of A and switches the toggle switch into the A off state and likewise, if S is expressed it turns the toggle switch on. However, as the implementation of a T latch with two auxiliary genes might have other pleiotropic effects, we reasoned that the introduction of a delay in the heterodimer-mediated autorepression of genes A and B could also greatly enhance the switching behavior between the two states. On a molecular level, such a delay can be realized by overlapping binding sites O_{TA} and O_{TB} for TA and TB, respectively, such that the heterodimers can only bind their target sites in a mutually exclusive manner. For strong binding sites with off-rates on the order of 10 minutes and longer, the heterodimer of the 'old' state, say TA, blocks the heterodimer of the 'new' state, say TB, long enough to promote a significant increase of the proteins in the new state (B high). If the toggle signal is removed at the right time, i.e., just when the new state is reached, our analysis showed that the circuit does indeed memorize the new state in a satisfactory manner (Fig. 3.7C). However, if the *toggle* signal is applied continuously, the circuit starts to oscillate between the two states with a period determined by the off-rate of the overlapping heterodimer binding sites (Fig. 3.7C). In digital electronics this oscillatory "race"-condition is typically omitted by introducing an external clock signal, which would in our case be akin to a toggle signal of just the right duration. As such an exact temporal control over genetic circuits is often hard to achieve, we will seek in the following for a circuit design which intrinsically accommodates for the inherent delays of its genes and is insensitive to the precise timing of its input signals.

J-K and Master-Slave latch. The J-K latch augments the *toggle* transition implemented in the T latch by providing the ability to control the internal state by the external input signals J and K, which can independently trigger the *set* and *reset* transitions, respectively (Fig. 3.6D). However, the J-K latch does not resolve the basic problem of the continuous oscillations in the presence of the *toggle* signal ($J = K = 1$) where it functions analogous to the T latch. Interestingly, in digital circuits that lack a global clock signal, that is, in asynchronous circuits, frequently an extension of the J-K latch – the "Master-Slave" latch – is used instead. The

central idea is to prevent continuous oscillations by "locking" the conventional J-K latch after half a period, thereby ensuring that the circuit's state gets uncoupled from the duration of the *toggle* signal (provided it is long enough). In digital electronics this is achieved by cross-coupling of a master and a slave latch via feedback regulation.

In the context of our genetic J-K latch this extension to a master-slave latch is achieved by additional feedback on the expression of the genes in the regulatory front end (Fig. 3.6E), such that (i) KA represses gene J and JB represses gene K and such that (ii) genes J and K repress each other to form an additional toggle switch. Qualitatively, if A is 'ON' initially, a *toggle* signal (here manifested in an elevated basal expression rate of genes J and K) leads to formation of KA, which represses both the synthesis of A as well as the synthesis of J. This switches the central toggle switch into the state B 'ON' and at the same time forces the toggle switch of the regulatory front end into the state K 'ON'. This state is then stably maintained for many cell generations, as verified by a quantitative model with physiological parameter values, see Fig. 3.7D for a dynamical test of all state transitions.

3.4. Significance of Sequential Logic in Natural Gene Circuits

Besides its potential applications in synthetic biology and engineering, some of the schemes of sequential logic described here might also be realized in naturally evolved gene networks. An important example in which cellular memory plays a pivotal role is the development of multicellular organisms, where choices in cell fate need to be inherited and maintained by their progeny throughout the lifetime of the organism. On a molecular level, different cell fates are often reflected by the multistability of the underlying regulatory network. For instance, the maturation decision of *Xenoopus* oocytes is based on a bistable phosphorylation cascade [195] and also the osteogenic differentiation of human mesenchymal progenitor cells seems to rely on a bistable switching mechanism [189]. Of particular interest are situations in which spatiotemporal signals (morphogens) lead to cellular differentiation patterns and thereby allow cells to transform transient signals received at a specific time during development and at a specific place in the tissue into a distinct cell state. In the following I will focus a well-studied example and scrutinize the significance of sequential transcription logic in this circuit.

Somitogenesis in Vertebrates

Vertebrate segments are formed during early embryogenesis, when vertebrae precursors, called somites, bud off in a rhythmic fashion from the anterior part of the presomitic mesoderm (PSM) [65], see Fig. 3.8A The periodic formation of somites is proposed to be controlled by the interaction of time-periodic gene expression of a somitic factor (*green stripes*) – the segmentation clock – with a morphogen gradient (*red line*) – the determination front – that moves along with embryonic elongation in posterior direction [34, 42, 54, 74]. At the determination front the oscillations of the somitic factor come to rest, and as it sweeps along the embryo it converts the periodic signal of the segmentation clock into a repetitive series of somites. From a operational point of view, this 'clock and wavefront' mechanism for vertebrate segmentation is akin to the behavior of a T latch. As illustrated in Fig. 3.8B, the morphogen gradient can be interpreted as the *toggle* signal (*red lines*), whereas expression of the somitic factor corresponds to the internal state of the T latch (*green lines*): In the posterior part of the PSM the *toggle* signal is 'ON', providing the command for cells to oscillate between two alternative expression states of the somitic factor. The determination front is

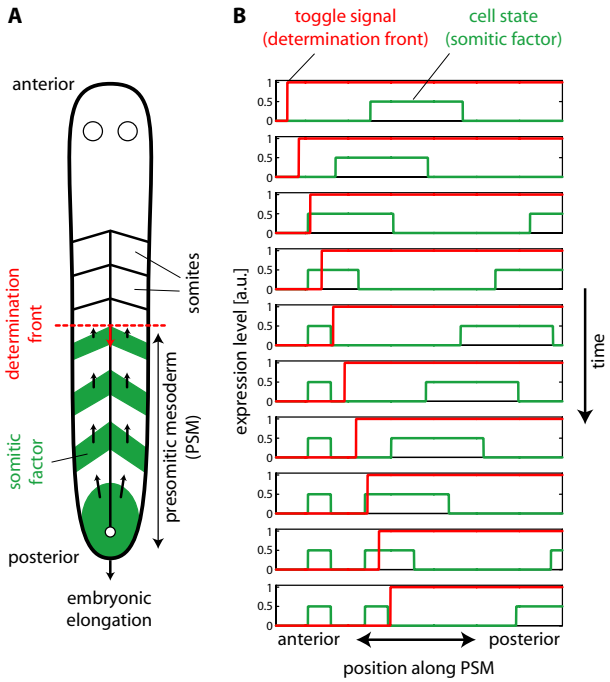


Figure 3.8.: (**A**) Clock and wavefront mechanism for somitogenesis in vertrebrate embryos (reproduced from Herrgen et al. [74]). (**B**) Possible mechanism of somite formation by spatiotemporal variations of the input signals to a D latch. For all details please refer to the main text.

reflected by the point at which the *toggle* command switches 'OFF'. As it progresses posteriorly along the PSM, cells stop oscillating and 'memorize' the current state of their oscillatory cycle. Interestingly, a similar pattern of repetitive stripes could also be generated by another sequential logic element, the D latch (see Fig. 3.6B). Within this hypothetical scenario, the *data* signal of the D latch would need to be coupled to an external clock, which provides an oscillatory input to all cells in the PSM. The sweeping maturation front would correspond to the *enable* signal, determining when cells should store the current level of the input signal in an independent memory element. As a consequence, the clock – reflected in the oscillatory data signal – would be required to continue ticking even in cells committed to their somitic cell fate.

On a molecular level, however, the realization of the clock and wavefront mechanism by a D latch seems not very likely. While a system of opposing gradients between the signaling molecules FGF (fibroblast growth factor)/Wnt and retinoic acid could indeed serve as a traveling maturation front in both models [42], there is currently no evidence for genes that continue cycling in somitic cells. In contrast, in the PSM many cyclic genes, such as the transcription factor HES1 (hairy and enhancer of split 1) [108], are known and it was shown that the Delta-Notch signal transduction pathway is involved in delayed coupling of neighboring cellular oscillators, leading to a wave-like expansion of the clock signal [74]. However, the precise interplay between the clock and the wavefront as well as the function of the segmentation clock is still a matter of ongoing debate. Hence, it seems that it is currently too speculative to seek for molecular components of a T latch. Nonetheless, I want to stress that the logic of the clock and wavefront mechanism very closely related to the sequential

logic operation performed by a T latch, and it might be stimulating to deeper analyze the segmentation clock from that perspective.

3.5. Conclusion

In this chapter a general perspective on the schemes of sequential logic in gene regulation was provided and a comprehensive classification of the building blocks for sequential logic design was given. Our analysis showed that all basic sequential logic devices fall into surprisingly few functional families and it was scrutinized how the molecular repertoire of bacterial gene regulation can be leveraged to find compact and robust genetic implementations. In particular, it was shown that protein heterodimerization provides a rapid and efficient realization of molecular AND gate logic, which can in turn be used as transcriptional control elements that selectively bind to programmable target sequences. While the proposed versions of sequential logic elements are guided by rational design and are thus intrinsically limited to a very narrow window of the possible design space, I feel that it might nevertheless be stimulating to analyze naturally occurring gene networks with memory from the angle of sequential logic design. Especially in eukaryotic cells, in which enhanceosomes take the role of large, multidimensional logic gates [7], one might suspect that similar schemes of sequential logic design could be realized.

In synthetic biology the biological implementation of these new devices might foster exciting applications, as they greatly expand the functional spectrum of available building blocks. For instance, the implementation of a conditional memory circuit in bacteria could serve as a whole-cell biosensor, which memorizes the state of a given environmental variable only under the condition of an externally supplied control signal, e.g., during a spatially and temporally restricted period of UV irradiation. Likewise, the implementation of bacterial counters could serve as sensor elements that count how often a specific chemical exceeds a predefined threshold value in a confined microenvironment.

4. Stochastic Timing of Gene Induction as a Regulation Strategy

Part of the work described in this chapter was [119] or will be published [120]. The full articles are attached in Appendices A.3 (Paper III) and A.4 (Paper IV), respectively.

The allocation of nutrients is a central task in microbial life. While bacteria often use glucose as a preferential carbon source, they also grow well on other, less ideal carbohydrates. To minimize the metabolic burden, most of the enzymatic machinery required for catabolism of these alternative growth substrates is only expressed when needed, that is, in the presence of the alternative substrate and in the absence of a better carbon source. Historically, Monod was the first to show in his seminal "diauxic shift" experiments in the 1940s, that when *E. coli* cells grow on a mixture of glucose and lactose, they first completely exhaust glucose from the medium and resume growth on lactose only after a characteristic lag phase [126]. From thereon, the study of the molecular origin of this "glucose effect" has shed light on many of the basic principles in bacterial gene regulation [145] and stimulated decades of research on the inducible carbon utilization systems.

A major finding along this way was made by Novick and Weiner [130], who discovered that the induction of the enzymes in the lactose utilization system (Lac system) displays an "all-or-none" behavior at intermediate levels of the non-metabolizable sugar analogue thiomethyl-β-D-galactoside (TMG). This means that a fraction of cells fully induced the Lac system (on-state) whereas the other didn't (off-state). Since the advent of modern, single-cell imaging techniques such bimodal expression patters have also been observed in the arabinose (Ara) system of *E. coli* [161] and in the galactose (*gal*) system of *S. cerevisiae* [2], suggesting that it is a common trait of the inducible carbon utilization systems. In fact, on a molecular level all of these systems share an autocatalytic positive feedback of the sugar on its own uptake proteins, which can give rise to bistable behavior in the presence of sufficient nonlinearities [51]. Meanwhile, the apparent bistability in these systems has attracted much attention and has become a paradigm for studying phenotypic heterogeneity in naturally occurring gene regulatory circuits [2, 48, 99, 134, 179].

The work presented in this chapter arose in close collaboration with Dr. Judith Megerle, Prof. Dr. Joachim Rädler (Experimental Biophysics, LMU Munich) and Noreen Walker (AMOLF, Amsterdam) and focusses on two previously unstudied aspects of the arabinose system in *E. coli*. First, we sought to examine the molecular mechanisms of single-cell switching between the off- and the on-state and *vice versa*, by the addition and removal of the inducer arabinose (Papers III & IV). In Paper III we studied the induction dynamics in the Ara system with single cell resolution and related the observed dynamics to a comprehensive mathematical model comprising the known biochemistry of the components in the Ara system. Here we quantitatively related the broadly heterogeneous timing in operon induction observed in our experiments to a wide distribution of arabinose uptake proteins at the time of inducer addition, and thereby provided a mechanistic explanation for the bimodal expression pattern

observed at sub-saturating inducer levels [161]. In Paper IV we mainly studied the down-regulation process upon sudden removal of external arabinose in single cells. Here our results show rapid and homogeneous switching from the on- to the off-state and shed light on a previously uncharacterized component of the Ara system, AraJ, which seems to function as an arabinose efflux valve. The second major aspect studied here focusses on the question whether the heterochronic gene induction is just an inevitable side-product of noisy gene expression in the Ara system, or whether the observed cell-to-cell variability confers any evolutionary advantage to a bacterial population. To approach this question, a mathematical framework based on the costs and benefits of gene expression is established, and it is shown that whenever gene induction is associated with a transient risk, but beneficial in the long-run, the stochastic timing in gene induction can maximize the reproductive success of a population in fluctuating environments.

The organization of this chapter is as follows: First, the molecular components of the arabinose system are introduced and their biochemical properties as required for our quantitative mathematical model are briefly reviewed (Section 4.1). Then the central results of Paper III on the heterogeneous timing in the induction behavior are presented (Section 4.2), followed by our findings of Paper IV on the down-regulation process and the role of AraJ in the Ara system (Section 4.3). Finally, a coarse-grained model for growth in fluctuating environments is presented, exemplifying the potential evolutionary advantage of heterogeneous timing in the induction of the carbon utilization systems (Section 4.4).

4.1. The Arabinose System of E. coli

The arabinose operon was the first system in which a positive regulatory mechanism was discovered. Unlike induction in the Lac system, which relies on deactivation of the lactose repressor LacI in the presence of lactose (double-negative control), the arabinose system is directly activated by the arabinose-responsive regulator AraC in the presence of arabinose. In the absence of arabinose, AraC also represses transcription efficiently by the formation of a DNA loop, leading to a large fold-change between basal (leaky) and maximal expression levels. This ability made the arabinose system central for biotechnological applications, and derivatives of its promoters serve as the working horses for heterologous gene expression [115]. Consequently, a lot of efforts have been undertaken to reduce some of the sources for heterogeneity that are described in this thesis [88–90, 127], and our results on the underlying mechanisms should prove useful for the rational design of new expression systems with predictable behavior.

Genes and Qualitative Behavior

The core of the arabinose utilization system of *E. coli* consists of the genes *araE* and *araFGH*, the products of which mediate arabinose uptake, and the products of the *araBAD* genes, catalyzing the three-step process of converting arabinose into D-xylose-5-phosphate, which ultimately feeds in the central metabolism of the cell (Fig. 4.1C,D). In addition, *araC* codes for the dual transcriptional regulator AraC, which – in the presence of arabinose – activates transcription of the promoters P_E, P_{FGH}, P_{BAD} and P_J, the promoter of *araJ*, a gene of yet unknown function [148, 156]. In the absence of arabinose, AraC also represses transcription from P_C and P_{BAD} by the formation of a DNA loop (Fig. 4.1B). This regulatory versatility of AraC is reflected in two alternative enzyme conformations [156]: in the absence of arabinose,

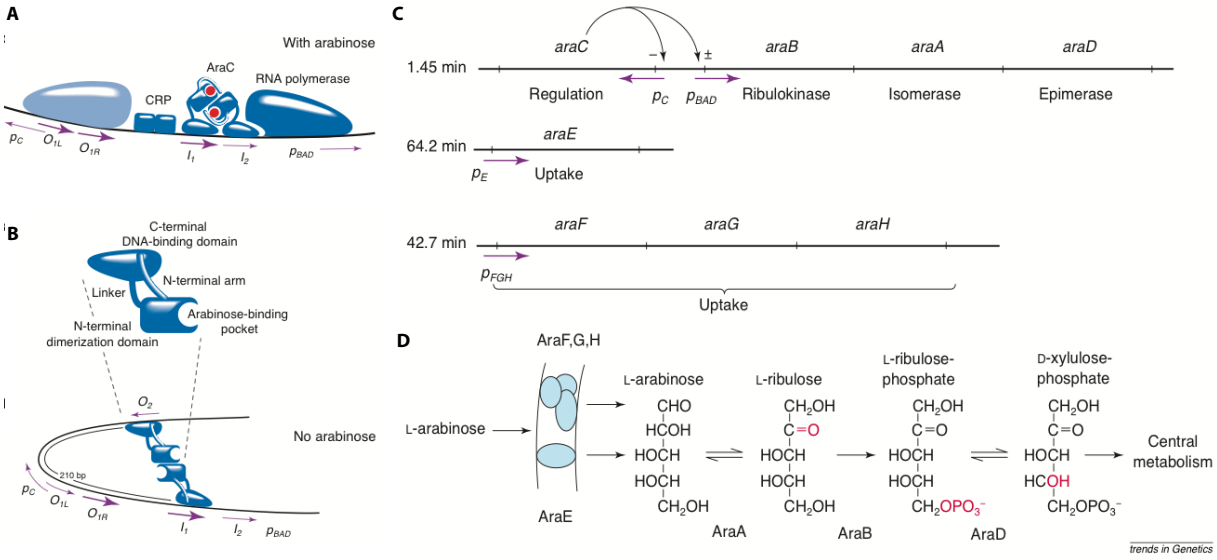


Figure 4.1.: Regulation (**A,B**), genes (**C**) and catabolic steps (**D**) in the arabinose utilization system of *E. coli*.[1] For all details please refer to the description in the main text.

the C-terminal and N-terminal subdomains of AraC are joined by a relatively rigid "linker arm", leading to an upright conformation in which the formation of the DNA loop by an AraC dimer is favored (Fig. 4.1B). Upon arabinose binding to AraC, the linker is released and the protein switches to a flexible conformation in which two adjacent binding sites upstream of its target promoters can be bound by an AraC dimer, where they recruit RNA polymerase and thereby activate transcription (Fig. 4.1A).

Qualitatively, the system functions as follows (Fig. 4.2): in the absence of external arabinose, cells express the genes in the Ara system at a basal level. Especially in the case of the arabinose uptake proteins this is of central importance, since arabinose is detected by import and internal sensing via AraC, and thus arabinose uptake proteins act as transporters and sensors alike [173]. Consequently, the promoters P_E and P_{FGH} also lack the upstream binding site required for DNA looping and display a significantly higher basal (uninduced) expression level than the P_{BAD} promoter [93]. As soon as the internal arabinose concentration reaches a certain threshold the promoters in the arabinose system are activated, and both the synthesis of further arabinose uptake proteins, as well the synthesis of the catabolic enzymes is stimulated. For the P_{BAD} promoter, the quantitative input-output characteristics suggest a cubic dependence on the internal arabinose concentration with an internal arabinose "activation threshold" of 50 μM. Effectively, this up-regulation closes both a positive feedback loop (uptake proteins) as well as a negative feedback loop (catabolic enzymes) on the internal arabinose level (Fig. 4.2). Since mutants deficient in arabinose catabolism were used throughout our studies, both the induction as well as the down-regulation dynamics upon changes in environmental arabinose are expected to rely critically on arabinose transport into (and possibly also out of) the cell.

[1] Reprinted from Trends in Genetics, Volume 16, Robert Schleif, Regulation of the L-arabinose operon of *Escherichia coli*, Pages 559-565, Copyright (2000), with permission from Elsevier.

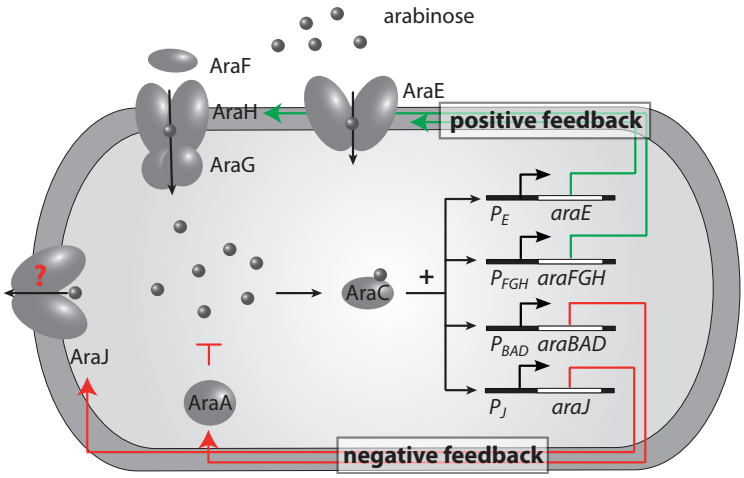


Figure 4.2.: Wiring diagram of the arabinose utilization system. Arabinose is imported via the arabinose transporters AraE and AraFGH. If the intracellular arabinose level is sufficiently large, arabinose binds the transcriptional regulator AraC. This complex activates the promoters P_E, P_{FGH}, P_{BAD} and P_J, driving expression of *araE*, *araFGH*, *araBAD* and *araJ*, respectively. The latter two operons encode genes for arabinose catabolism (*araBAD*) and a putative arabinose efflux pump (*araJ*). Arrows indicate arabinose transport and positive regulation, whereas the T-shaped arrow indicates arabinose catabolism.

Arabinose Transport

Arabinose uptake is catalyzed by the high affinity, low capacity transporter AraFGH, as well as by the low affinity, high capacity transporter AraE [75]. The AraFGH transporter is a member of the ATP Binding Cassette (ABC) transporter superfamily, featuring the periplasmic binding protein AraF, the membrane component AraH and the ATP-binding component AraG [78] (Fig. 4.2). It has an apparent affinity for arabinose of 1 μM [75] and relies on ATP hydrolysis during transport [36]. In contrast, AraE is an arabinose/H^+ symporter with a low affinity of $140 - 320\,\mu$M [36]. However, comparison of arabinose uptake in wild-type strains with *araE* and *araFGH* deletion strains revealed that the two transporters do not operate independently [36] and that instead, arabinose transport was best described by a single Michaelis-Menten function. The origin of this synergistic interaction between the two transport systems is not resolved. Together the two systems accumulate arabinose more than 300-fold at low (20 μM) and about 30-fold at high (1 mM) external arabinose concentrations [131]. In Paper III we focus on the influence of transporter heterogeneity on the induction behavior of the Ara system and demonstrate that it directly relates to an experimentally observed heterogeneous time delay in gene induction.

In addition to arabinose uptake, there is also evidence that arabinose is rapidly exported form the cell in a mutant deficient in arabinose catabolism (*araA*$^-$). In a seminal study, Novotny and Englesberg [131] showed that arabinose export follows first order kinetics up to high internal arabinose concentrations with a rate constant of $k = 5.2\,\mathrm{min}^{-1}$ at 37°C. On a molecular level it seems plausible that the product of the fourth arabinose-inducible gene, *araJ*, is involved in the observed export, since sequence similarity with drug efflux proteins of the major facilitator superfamily [135] suggests that it may function as a sugar efflux system [87]. Hence, while there is an apparent lack of function of *araJ* for the induction of the *araBAD* operon [148], it might speed up the transition from the on- to the off-state upon

removal of external arabinose.

Interestingly, transcriptional activation of *ydeA* has been shown to interfere with arabinose accumulation and induction of the P_{BAD} promoter [20, 27]. *ydeA* also belongs to the major facilitator superfamily and shows sequence similarity to *araJ* (24% identity) [20]. However, *ydeA* is only expressed at extremely low levels in exponentially growing wild-type cells and is not induced by arabinose [20]. Hence, while *ydeA* appears non-essential under the standard laboratory conditions tested, it has been speculated that YdeA, when expressed, and/or AraJ could promote the export of arabinose structural analogues that can be imported but not completely metabolized [20]. The fact that *araJ* is induced severalfold in the presence of arabinose [73] makes it tempting to speculate that it plays a central role in the physiological response of *E. coli* towards arabinose and its close chemical analogues. In Paper IV we focus on the influence of arabinose efflux on the on- and off-switching dynamics and show that *araJ* has a clear phenotype during off-switching of the Ara system.

4.2. Paper III: Timing and Dynamics of Single Cell Gene Expression in the Arabinose Utilization System

In the paper 'Timing and dynamics of single cell gene expression in the arabinose utilization system', *Biophys. J.* **95**, 2103-2115 (2008), by Judith A. Megerle*, Georg Fritz*, Ulrich Gerland, Kirsten Jung, and Joachim O. Rädler (*equal contribution) we presented a quantitative single-cell analysis of gene induction in the arabinose system of *E. coli*. To the best of our knowledge, this work reports for the first time the experimental single-cell expression dynamics associated with the decision process that triggers switching from the off- to the on-state. The reported experiments were performed in controlled, microfluidic environments, which permit the addition of arabinose *in situ* and allowed us to follow the expression dynamics in single cells, rather than measuring only their time-dependent population distributions. To study the switching dynamics under the microscope, an *E. coli* strain deficient in arabinose catabolism ($araBAD^-$) was used and was transformed with the reporter plasmid pBAD24-GFP, in which the rapidly maturing GFP variant *gfpmut3* is under the control of the P_{BAD} promoter.

Interestingly, at intermediate inducer levels we found a broadly heterogeneous time delay in the onset of gene expression, whereas at high inducer concentrations cells responded homogeneously and with minimal delay. With the help of a separate measurement of the cell-to-cell variability in the GFP maturation time, we extracted the distribution of lag times intrinsic to the arabinose uptake process. This was achieved by establishing a deterministic GFP expression model within single cells and through fitting the model to the experimental fluorescence trajectories with only two key parameters: the maximal GFP expression rate and the delay time after which the P_{BAD} promoter gets turned on. From our analysis we found that the distribution of maximal GFP expression rates is independent of the inducing arabinose concentration, whereas both the average as well as the width of the delay time distribution scaled inversely with the external arabinose level.

To describe our experimental lag time distributions, we developed a theoretical framework which supports the conclusion that the observed heterogeneity arises from cell-to-cell variations in the number of arabinose uptake proteins at the time of sugar addition. In this model for the induction process we did not consider potential effects of arabinose efflux, Instead, we made the parsimonious assumption that the time delay in the induction process results

from slow arabinose uptake alone - an hypothesis which seemed compatible with the range of arabinose uptake velocities by AraE and AraFGH, as inferred from experimental transport studies [36]. Assuming that the number of transporters does not change significantly during the induction process (diabatic limit), the model then predicted that the time delay, τ_d, until a certain internal arabinose threshold required for promoter activation is reached, scales inversely with the number of transporters, n, at the time of inducer addition ($t = 0$). The steady-state distribution $P(n)$ of arabinose transporters at $t = 0$ is the result of a bursty, basal activity of the promoters P_E and P_{FGH}, and takes for a single uptake protein species the form of a negative binomial distribution [13],

$$P(n) \approx \left(\frac{1}{1+b}\right)^{\mu} \left(\frac{b}{1+b}\right)^{n} \begin{pmatrix} \mu+n-1 \\ n \end{pmatrix}, \tag{4.1}$$

where the "burst size" b is the ratio of the translation rate and the mRNA degradation rate and corresponds to the typical number of proteins produced from a single mRNA [178]. Likewise, μ is the ratio of the basal transcription rate and the protein dilution rate and can be interpreted as a dimensionless "burst frequency" (the number of bursts within the lifetime of a protein). Together with the relation $\tau_d \sim 1/n$, the distribution of delay times $Q(\tau_d)$ can be obtained from the transformation rule $Q(\tau_d) = \left|\frac{dn(\tau_d)}{d\tau_d}\right| P(n)$, yielding

$$Q(\tau_D) \approx \frac{\tau_0}{\tau_D^2} \left(\frac{1}{1+b}\right)^{\mu} \left(\frac{b}{1+b}\right)^{\tau_0/\tau_D} \frac{\Gamma(\tau_0/\tau_D+\mu)}{\Gamma(\tau_0/\tau_D+1)\Gamma(\mu)}, \tag{4.2}$$

where $\Gamma(x)$ is the Gamma function and τ_0 is a parameter determined by the external arabinose concentration.

From pairwise Kolmogorov-Smirnov tests at each external arabinose concentration, we found that this theory consistently fits the shape of the experimental delay time distributions for various inducer concentrations. Hence, our data support a previous conjecture by Siegele and Hu [161], according to which the delay time distribution is causally linked to the distribution of uptake proteins in the absence of the inducer. However, while our model only accounts for the dynamics of arabinose uptake and thus – according to Occam's razor – represents the most parsimonious explanation of the induction behavior of the arabinose system, we suspected that under altered experimental conditions also arabinose export might play a critical role. Therefore, in Paper IV we analyzed the influence of the putative arabinose efflux protein *AraJ* on both the switching from the off- to the on-state, as well as the switching from the on- to the off- state of the Ara system, see Section 4.3.

Additionally, our characterization of the heterogeneous timing in the switching between alternative phenotypic states forms the basis to further explore possible links between the single-cell dynamics and biological function of these systems. For instance, in fluctuating environments, heterogeneous switching dynamics may prevent costly synthesis of specialized enzymes in all cells of a colony, when the enzyme substrate is available only for short periods. A first attempt to make such a link between heterogeneous timing and physiological function in the Ara system is the subject of Section 4.4.

4.3. Paper IV: Quantitative Characterization of Single Cell Switching Dynamics in the Arabinose Utilization System

In the submitted manuscript 'Quantitative characterization of single cell switching dynamics in the arabinose utilization system' by J. A. Megerle*, G. Fritz*, S. A. Westermayer, D. Brick, R. Heermann, K. Jung, J. O. Rädler, and U. Gerland (*equal contribution) we analyzed the role of the putative arabinose exporter AraJ in the switching dynamics of the arabinose system. To test whether AraJ has indeed arabinose export activity *in vivo* as predicted (see Section 4.1), we studied the down-regulation of the P_{BAD} promoter upon sudden removal of external arabinose. By conducting our experiments in cells deficient of arabinose catabolism (*araBAD*$^-$), we sought to eliminate other potential arabinose sinks and reasoned that the dynamics of internal arabinose (controlling the P_{BAD} promoter activity) should be mainly affected by arabinose efflux. To single out the effect of AraJ on arabinose efflux, we compared the switching dynamics in an *araJ*$^+$ and in an *araJ*$^-$ strain. In the *araJ*$^+$ strain, our single cell experiments revealed that fluorescence production from P_{BAD} ceases 20 - 30 min after removal of external arabinose. In contrast, the *araJ*$^-$ strain continued fluorescence production more than 50 min after arabinose removal, indicating that *araJ* clearly speeds up the transition from the on- to the off-state in the arabinose system. Comparison with a refined quantitative mathematical model revealed, that the rapid timescale of arabinose efflux ($k^{-1} < 1$ min) reported by Novotny and Englesberg [131] consistently accounts for the observed fluorescence trajectories.

To further corroborate our model, we studied the effect of rapid arabinose efflux on the heterogeneous timing in gene induction described above. Here it turned out, that the rapid export can be balanced by a faster import rate (still within the range of arabinose uptake velocities inferred from experimental transport studies [36]), leading to a similar timing heterogeneity as before. Within our refined model, the mechanism behind this delay is as follows: Upon arabinose addition, internal arabinose equilibrates rapidly (< 1 min) to a steady state value, a^*_{in}, that is determined by the ratio of arabinose in- and efflux rates, i.e., $a^*_{in} \sim n/k$. Thus, cell-to-cell variability in the initial number of arabinose uptake proteins, n, directly affects the initial arabinose level within each cell. Since internal arabinose activates expression of the genes in the Ara system, this leads to a cell-to-cell variability in the initial rates of gene expression. Due to the positive feedback of the uptake proteins on their own expression the cellular level of transporters increases, and meanwhile also the internal arabinose concentration follows immediately after a change in transporter level (adiabatic limit). As a result, the promoter activity increases to its maximal value on a timescale determined by the (slow) kinetics of AraE expression. Accordingly, cells with a high initial number of uptake proteins fully activate the Ara system earlier than cells with a low initial number of uptake proteins. Interestingly, the delay time until promoter activation scales for typical levels of basal transporter expression (∼20-200 proteins) in good approximation inversely with the number of transporters ($\sim 1/n$) (Fig. 4.3A). Thus, the refined model of the Ara system presented here leads to a very similar heterochronic induction behavior than in our model without arabinose efflux (Fig. 4.3B,C), indicating that both models are quantitatively consistent with the experimental switching kinetics from the off- to the on-state of the Ara system.

However, besides accounting for the switching from the on- to the off-state in the Ara system, our refined model also predicted the behavior of a mutant with constitutively low transporter expression, that is, in a strain with a disrupted positive feedback loop of the

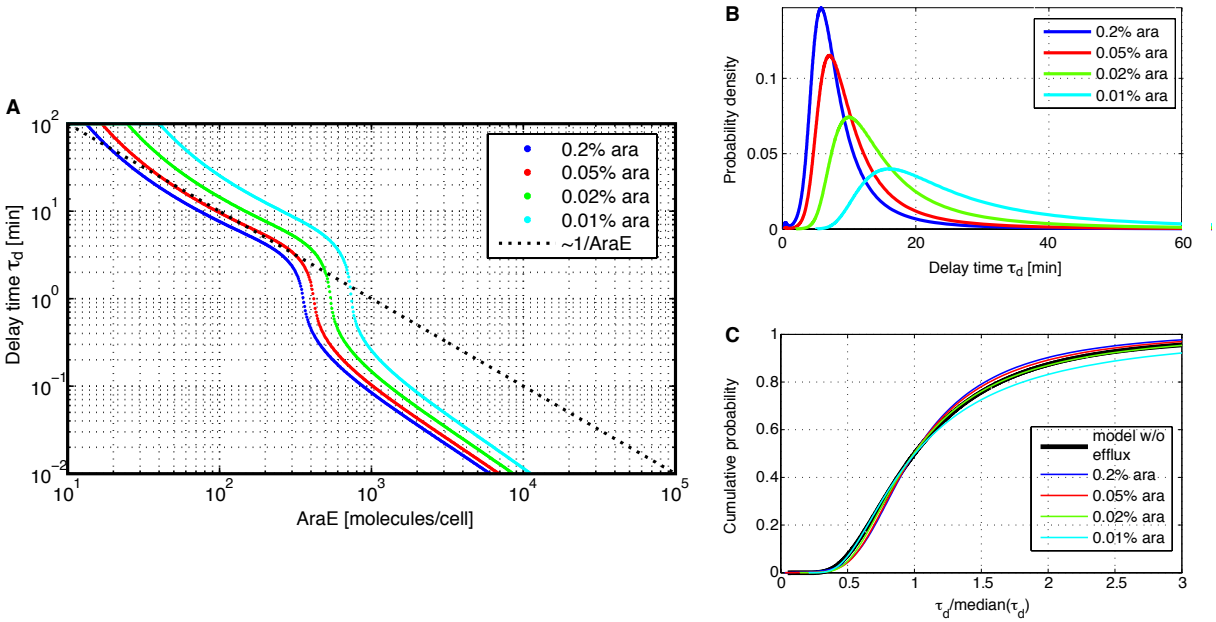


Figure 4.3.: (**A**) Theoretical delay time as a function of the initial number of transporter proteins AraE. The delay time was defined as the time until the PBAD promoter reached 95% of its maximal activity and obtained from our refined mathematical model, for details see Paper IV. Theoretical delay time distributions (**B**) and cumulative probabilities (**C**) at different external arabinose concentrations in our model with arabinose efflux, given a negative binomial distribution of initial uptake proteins. Due to the inverse scaling of the delay time with the number of AraE proteins over the relevant rage (**A**; 20-200 proteins), the delay time distributions obtained with the refined model (**C**; *colored lines*) only differ within experimental error from our previous results for the delay time distribution in a model without arabinose efflux (**C**; *black line*).

transporters on their own expression. In this strai all cells in the culture should be induced instantly, that is without heterogeneous delay. Their expression rate, however, should be sub-maximal and we expected a modulation of the average gene expression rate instead of a modulation in the delay times upon variation of the external arabinose concentration. Experimental verification of these predictions by recording the single cell expression dynamics in a mutant with P_{lac}-driven *araE* expression has put our refined model including arabinose efflux on firm grounds.

4.4. Regulated Bet-Hedging Through Stochastic Timing of Gene Induction

In the previous sections it was demonstrated that the arabinose system displays heterochronic gene induction upon inducer addition, which can be interpreted as a "responsive stochastic switching" strategy. In this regulation strategy, the detected environmental change in the sugar concentration affects the response of a single cell in two ways: (i) it triggers the switch to the 'ON state' after a stochastic delay, and (ii) the sugar concentration controls the shape of the probability distribution for the delay, such that the timing of gene induction is highly variable at low concentrations and nearly homogeneous at high concentrations. To address the

functional question, as to whether responsive stochastic switching confers an advantage over other possible regulation schemes, a coarse-grained mathematical model for bacterial growth in unpredictable fluctuating environments will be devised. An essential property of the model is a transient risk associated with induction of the system, arising from a cost-benefit analysis of gene expression where protein synthesis incurs an immediate cost and only a delayed enzymatic benefit [39, 40, 82]. Given such a finite "amortization time" and an unpredictably fluctuating environment, one can ask whether the observed responsive stochastic switching naturally arises as a regulated bet-hedging strategy.

4.4.1. Cost-Benefit Estimate of Gene Induction

In order to study the potential risks and advantages of heterogeneous timing in gene induction within a microbial population, let us first focus on the costs and benefits associated with gene expression within a single cell. By means of a simple mathematical quantitative model for the cellular energy balance, it will be shown now that there can be a transient phase during which cells invest more energy into the expression of the sugar utilization system than they harvest by sugar digestion. As a basis for the analysis, let us consider a scenario in which cells are suddenly shifted from a nutrient-free to a nutrient-containing environment. Before the shift, cells are in a non-growing (stationary) state in which they possess a pool of intracellular building blocks and energy equivalents, which are required for maintenance of their metabolic activity and for future gene expression.

Without loss of generality, we assume that in the pool of cellular energy is the limiting factor for cell growth and survival[1], and consider for simplicity the ATP level as the universal "energy currency" of the cell. Converting all other energy compounds into ATP equivalents, the total cost of synthesizing an average protein of 360 amino acids length is $\zeta = 1500$ ATP/protein [169]. Likewise, under ideal conditions the full digestion of glucose yields $\beta = 36 - 38$ ATP/glucose molecule [169]. However, the yield and is typically smaller for other carbon sources and suboptimal growth conditions. For instance, fermentation of glucose during anaerobic growth only yields $\beta = 2$ ATP/glucose molecule and other sugars provide even less energy. Throughout this section a value of $\beta = 1$ ATP/sugar molecule will be used for the energy content of an average sugar molecule under typically poor growth conditions and a value of $\zeta = 3000$ ATP/protein for the average expression cost of the sugar utilization operon[2].

Supposing that cells induce protein synthesis after a delay-time τ at rate κ and noting that cells do not grow until a certain energy threshold is reached, the number of transporters per cell increases linearly with time. Given that a single transporter imports sugar at an effective rate $v_S = v_{max}S/(K_m + S)$, with v_{max} being the maximal uptake rate, K_m the Michaelis constant and S the external sugar concentration, and assuming that internal sugar is instantaneously converted into energy, the rate of energy production is also a linear function of time, that is, $\dot{E}^+ = \beta v_S \kappa(t - \tau)$. Contrary, the rate of energy consumption is proportional to the rate of protein synthesis and hence constant in time, i.e., $\dot{E}^- = \zeta\kappa$. Together, the

[1] Our analysis is equally applicable under conditions in which one of the building blocks (amino acids, nucleotides, etc.) is limiting – with only a few modifications of the parameter values.

[2] In the arabinose system, sugar is imported either via the arabinose/H^+ symporter AraE (1 protein) or the ABC transporter AraFGH (3 proteins). Since sugar import is the bottleneck for energy production [92], we suppose that transporter expression poses the most significant cost to the cell and take an average value of 2 proteins required for sugar uptake.

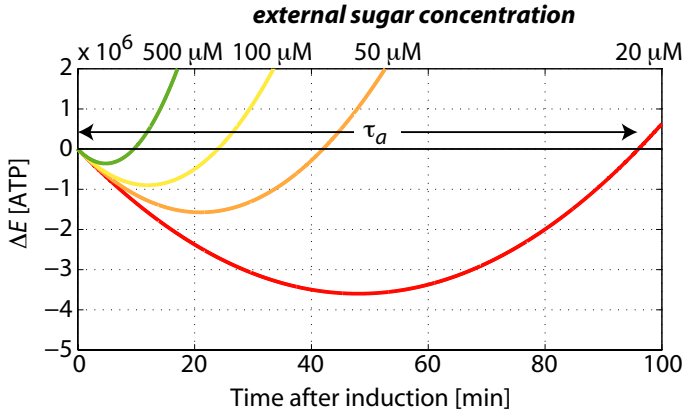


Figure 4.4.: Dynamics of internal energy upon gene induction at indicated external sugar concentrations. The amortization time τ_a is marked for 20 μM external sugar. The parameters are chosen similar to those of the arabinose system: $\kappa = 50$ proteins/min, $K_m = 300\,\mu$M and $v_{max} = 1000$ sugars/protein/min, $\beta = 1$ ATP/sugar, $\zeta = 3000$ ATP/protein.

change of the internal energy level, $\Delta E(t) \equiv E(t) - E(0)$, exhibits parabolic dynamics, that is,

$$\Delta E(t) = \beta v_S \kappa (t-\tau)^2/2 - \zeta\kappa(t-\tau)\,. \tag{4.3}$$

Taking the *in vivo* parameters of the arabinose utilization system, Fig. 4.4 shows the time-dependence of ΔE at different external sugar concentrations. Note that for all sugar concentrations there is a transient period during which the cells invest more energy than they gain at that time. Importantly, the duration τ_a of this amortization period scales inversely with the effective rate of sugar import and is hence a function of the external sugar concentration, that is,

$$\tau_a = \frac{2\zeta}{\beta v_S}\,. \tag{4.4}$$

At saturating external sugar levels, when all transporters operate at their maximal rate ($v_S = v_{max}$), the initial investment into a protein is amortized in about 6 minutes. However, it is known that both the arabinose as well as the lactose operon already respond to sugar concentrations in the micromolar regime [97, 130, 134, 155] and under such substrate-limiting conditions the situation can be vastly different: For instance, at an arabinose concentration of 20 μM the amortization time may be as long as 95 min (cf. Fig. 4.4A; *red curve*), which is comparable to typical lag-phase durations [104, 133].

Notably, the maximal energy deficit of about 4×10^6 ATP experienced for $S = 20\,\mu$M is on the order of the internal ATP level of *E. coli* in balanced growth: the maximal ATP level in balanced growth is about 10 mM, corresponding to $E_{max} \approx 5 \times 10^6$ ATP/cell [169], which serves us as an upper limit for the energy level in stationary phase. From this – admittedly crude – estimation it becomes apparent that the entire energy reservoir of a cell can become exhausted within a few successive nutrient pulses of low concentration and unfortunate duration.

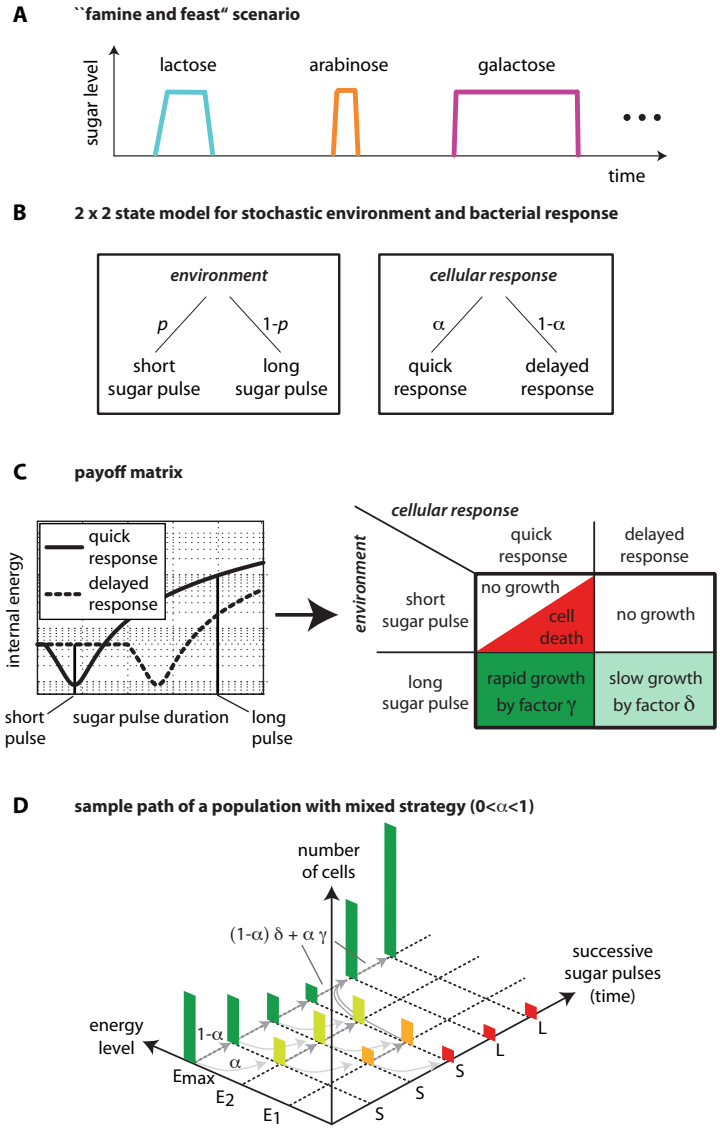


Figure 4.5.: Simple model for growth and survival in an unpredictable environment. (**A**) In a natural growth scenario periods of famine (no sugar) are interrupted by periods of feast (sugar available). (**B**) Simple binary model extracting the essential features of the famine and feast scenario. (**C**) The payoff matrix relates the change in internal energy to the growth factors μ (cf. Eq. (4.6)) for all combinations of pulse durations and cell responses: Upon long sugar pulses cells reproduce by a factor γ or δ, depending on whether they responded quickly or delayed, respectively. The energy deficit of a cell responding quickly in a short sugar pulse is only a fraction of the maximal internal energy, such that cells do not suffer immediately from short sugar pulses. However, cells in the lowest (critical) energy level die if they respond quickly in a short sugar pulse. (**D**) In a mixed strategy, only the quickly responding subpopulation (fraction α) looses energy upon short sugar pulses (S), whereas the energy level of the delayed subpopulation (fraction $1-\alpha$) remains unchanged. Cells reaching the lowest energy level cannot resume growth (=dead cells), whereas cells with intermediate energy levels do grow upon long sugar pulses and reach the highest energy level E_{max}. At the same time long sugar pulses (L) cause population growth by a factor $(1-\alpha)\delta + \alpha\gamma$.

4.4.2. Growth Optimal Enzyme Production in Fluctuating Environments

In natural environments of microorganisms the substrate abundances fluctuate with time and typically phases of exponential growth are interrupted by periods of stationary phase, during which growth is halted. In such a famine and feast scenario, the durations of nutrient abundance are controlled by external factors and can generally not be "anticipated" by the bacterial population. Likewise, also the type of nutrient in future nutrient pulses is barely predictable as it depends, e.g., on the nutritional behavior of the host organism or on stochastic nutrient flows in the soil. Based on these general considerations let us now focus on an idealized growth scenario in which cells experience a repetitive series of sugar pulses (Fig. 4.5A). Here the type of sugar is always different from the sugar in the previous pulse and can only be accessed through induction of the corresponding carbon utilization system. Hence, cells cannot profit from pre-induction of sugar utilization systems in consecutive pulses and it is plausible that cells show no correlation between the response-times in successive sugar pulses. Furthermore, we assume that there is neither a correlation between the energy level of a cell and its response-time nor between energy level and the rate of protein synthesis.

For simplicity let us consider a two-state environment and a binary (stochastic) response of the population (Fig. 4.5B): The environment produces short (S) and long (L) sugar pulses with probability p and $1-p$, respectively. The sugar concentration is assumed to be identical in all sugar pulses and is taken as a tunable parameter to our model. Importantly, we focus on a concentration regime in which cells experience a transient decrease in the internal energy level, such that the induction of the system is disadvantageous during the amortization period but advantageous in the long run. As cells cannot anticipate the duration of a sugar pulse at the time when the sugar appears, it is plausible that the population splits stochastically into a fraction of quickly and slowly responding cells with probabilities α and $1-\alpha$, respectively. Hence, the genotype is encoded in the probability α that determines the fractions of two phenotypically distinct subpopulations. This should not be confused with the evolutionary competition *between* different genotypes, e.g., quick vs. delayed genotype.

To quantify the fitness of the quick and delayed subpopulations with respect to a given sugar pulse duration, first focus on the associated change of internal energy within a single cell (Fig. 4.4). Here we assume that the duration of the short sugar pulse falls within the amortization period of the quickly responding strategy, whereas it still falls into the delay-time of the delayed subpopulation. Hence, the quick subpopulation suffers an energy deficit ($\Delta E = \varepsilon < 0$) from a short sugar pulse, while the delayed subpopulation stays inert ($\Delta E = 0$). Contrary, the long sugar pulse is chosen longer than the amortization period of the delayed subpopulation, such that both strategies experience an energy gain from a long pulse. However, due to the different induction times cells with a quick response foster much more energy than cells with a delayed response.

In the next step we relate this change in internal energy to cell growth. To this end, it is assumed that cells keep memory of their internal energy level from sugar pulse to sugar pulse and start growing as soon as they reach the maximal internal energy level E_{max} of *E. coli* in balanced growth (see Section 4.4.1). Contrary, cell death occurs as soon as their energy pool gets depleted. For long sugar pulses, suppose that the energy gain is for both response strategies much larger than E_{max}, such that all cells are reset to the highest energy level and use the surplus of energy for reproduction. As this surplus is different for both response strategies, the quickly responding subpopulation has a higher growth factor, that is, a higher number of offspring per sugar pulse, than the delayed subpopulation, i.e., $\gamma > \delta$

(cf. Fig. 4.5C). For short sugar pulses, cells either stay inert or experience an energy deficit ε, such that cells can be in one out of q discrete energy levels $\{E_{max}, E_{max}-\varepsilon, E_{max}-2\varepsilon, ..., E_{max}-(q-1)\varepsilon\}$ before it dies (Fig. 4.5D). Here the number of energy levels q is a function of the external sugar concentration, as it determines the magnitude of ε. For instance, a sugar concentration of $S = 50\,\mu$M corresponds to $q = 3$, meaning that cells survive a sequence of two short sugar pulses, but die in the third one due to energy depletion.

Long-term population growth rate

Generally, the long-term population growth rate λ can be defined through

$$\lambda \;=\; \lim_{T\to\infty} \frac{1}{T} \log_2 \frac{N_T}{N_0}\,, \tag{4.5}$$

where time T is a discrete measure for the number of sugar pulses and the total population size N_T after sugar pulse T is described by a Markov chain

$$N_T = \mu_T N_{T-1} = N_0 \prod_{i=1}^{T} \mu_i\,, \tag{4.6}$$

with N_0 being the initial population size. Here the μ_i are the (stochastic) factors by which the population grows in sugar pulse i. They depend both on the realization of the environment (short or long sugar pulse, abbreviated as S or L pulse, respectively) and on the fraction α of cells that responds with the quick strategy (see below). By combining Eqs. (4.5) and (4.6) the long-term growth rate can be expressed as an average over all logarithmic growth factors, that is,

$$\lambda = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} \log_2 \mu_i \equiv \overline{\log_2 \mu}\,. \tag{4.7}$$

However, since cells contain a certain memory about their history in form of their internal energy level, the growth factors not only depend on the duration of the current sugar pulse but rather on the stochastic sequence of all previous pulse durations: The more short pulses arrive in a row ("S-chain") the more dangerous the environment gets for the population. For instance, the sequence SLSLSL is harmless with respect to extinction of a population, since long sugar pulses reset all cells to the highest energy level in every second sugar pulse, whereas the sequence LLLSSS causes extinction of cells that respond quickly three time in a row during the last three pulses.

Consequently, the average of the logarithmic growth factors in Eq. (4.7) has to be taken over all possible sequences of sugar pulse durations. The analysis simplifies, however, by noting that a long sugar pulse always resets the whole population to the highest energy level and thereby "erases" the memory of all cells. Hence, during long sugar pulses which occur with probability $(1-p)$ the population always grows by a factor $\mu_{\mathrm{L}}(\alpha) = (1-\alpha)\delta + \alpha\gamma$. In contrast, the survival of a short sugar pulse depends on the history of previous pulse durations and one has to to average over all sequences of S-chains. The probability to find the sequence $\mathrm{LS}^n\mathrm{L}$, i.e., the probability to have exactly n short pulses in a row, is given by $(1-p)p^n(1-p)$. The associated growth (or better: survival) factor is determined by the fraction of cells which responded quickly at most $(q-1)$ times within the S-chain of length n, i.e.,

$$\mu_{\mathrm{S}^n}(\alpha) \;=\; \sum_{k=0}^{q-1} \binom{n}{k} \alpha^k (1-\alpha)^{n-k}\,. \tag{4.8}$$

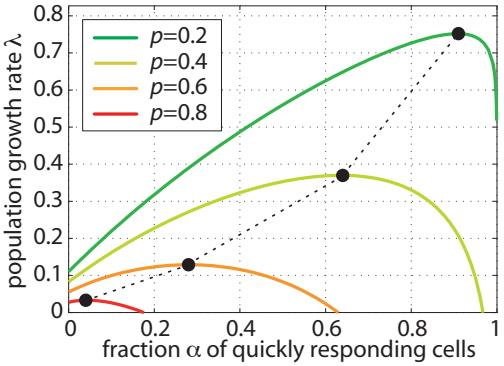


Figure 4.6.: The population growth rate λ is maximal for mixed response strategies ($0 < \alpha < 1$). The model parameters are $q = 2$ internal energy states and $\gamma/\delta = 2$, with $\gamma = 2.2$ and $\delta = 1.1$.

With this, the long-term population growth rate in Eq. (4.7) takes the form

$$\lambda(\alpha) \quad = \quad (1-p)\log_2 \mu_{\mathrm{L}}(\alpha) + (1-p)^2 \sum_{n=3}^{\infty} p^n \log_2 \mu_{\mathrm{S}^n}(\alpha). \tag{4.9}$$

Optimal Response Strategies

Next, we optimize the response-strategy with respect to the long-term population growth rate λ, by varying the fraction of quickly responding cells, α, from 0 to 1. Strikingly, the population growth rate is a concave function and displays a maximum at intermediate values of α (Fig. 4.6): For a given environment, defined through its probability p to provide short sugar pulses, the population grows at a basal rate $\lambda = (1-p)\log_2 \delta$ if all cells play the conservative strategy ($\alpha = 0$). For increasing α the growth advantage of quickly responding cells outweighs the risk of extinction, since a substantial fraction of cells with delayed response buffer against such fatal events. At high α however, this risk can no longer be compensated for, the population growth rate decreases and even becomes negative, indicating the extinction of the whole population in the long run. Clearly, the higher the probability of short sugar pulses (high p) the more dangerous is the environment for the population, thus driving the optimal strategy towards a more conservative response (α low) (Fig. 4.6; *red line*). Contrary, the more long sugar pulses appear (low p), the more beneficial it is to have a large fraction of quickly responding cells (Fig. 4.6; *green line*).

These results indicate that a heterogeneous strategy, reflected in a population of mixed phenotypes, can indeed be superior to a homogeneous strategy. Next, let us scrutinize how the particular choice of model parameters affect this result and under which conditions a homogeneous response might be favorable. To this end, we test how the growth rates γ and δ, as well as the probability p of a short sugar pulse affect the optimal response strategy. This can be done by numerically finding the roots of

$$\frac{\partial \lambda}{\partial \alpha} \quad = \quad (1-p)\frac{\frac{\gamma}{\delta} - 1}{\alpha\frac{\gamma}{\delta} + (1-\alpha)} + \alpha^{q-1} g_q(\alpha, n)\,, \tag{4.10}$$

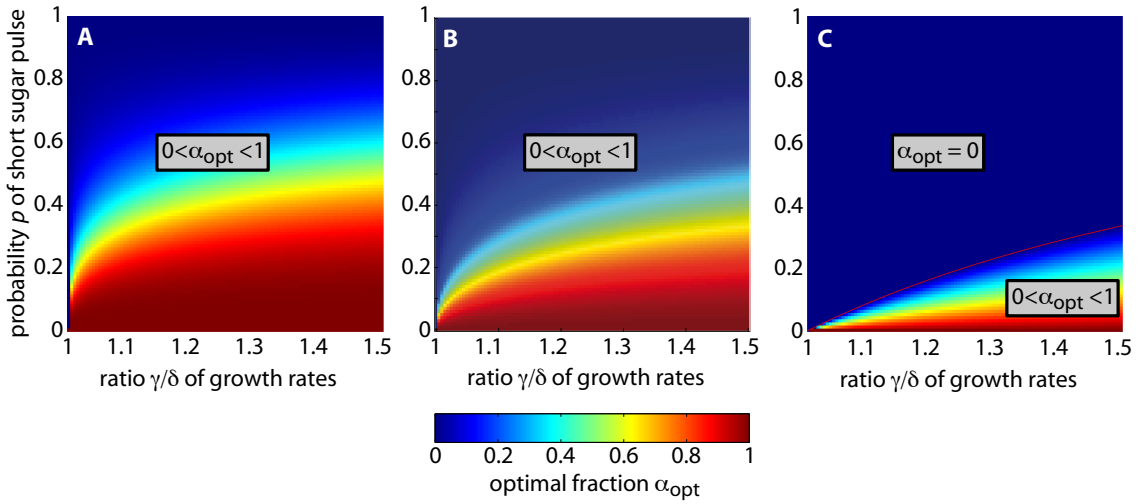


Figure 4.7.: Optimal response strategies α_{opt} as a function of p and γ/δ. (**A**) 3 internal states, (**B**) 2 internal states and (**C**) 1 internal state. The red line indicates the boundary between a heterogeneous ($0 < \alpha_{opt} < 1$) and a homogeneous ($\alpha_{opt} = 0$) response.

where $g_q(\alpha, n)$ is a function that is non-singular and continuous at $\alpha = 0$, absorbing the infinite sum resulting from the second term in Eq. (4.9). Note that the optimal response not depends the individual growth rates of the quick and delayed subpopulations, but only on their ratio γ/δ. However, in order to identify the region in which a homogeneous response ($\alpha = 0$ or $\alpha = 1$) is favorable over a heterogeneous response ($0 < \alpha < 1$), a different approach is needed, since for very small α the numerical solution cannot distinguish between $\alpha = 0$ (exact) and $\alpha > 0$, but very small.

To this end we exploit that $\lambda(\alpha)$ is concave, i.e., $\frac{\partial^2 \lambda}{\partial \alpha^2} < 0$, and hence, if $\frac{\partial \lambda}{\partial \alpha} \mid_{\alpha=0} > 0$ the maximum of $\lambda(\alpha)$ is located at $\alpha_{opt} > 0$, whereas for $\frac{\partial \lambda}{\partial \alpha} \mid_{\alpha=0} < 0$ the optimal strategy is located at $\alpha_{opt} = 0$, since negative values of α are not allowed. Inserting $\alpha = 0$ into Eq. (4.10) and noting that $\gamma/\delta > 1$ (quick cells grow faster than delayed cells), yields

$$\frac{\partial \lambda}{\partial \alpha} \mid_{\alpha=0} = (1-p)(\frac{\gamma}{\delta} - 1) > 0 \qquad \forall p \neq 1\,. \tag{4.11}$$

Hence, cells should only respond delayed in every pulse ($\alpha_{opt} = 0$), if the environment generates short sugar pulses with certainty ($p = 1$). Intuitively one would suspect that such a conservative strategy is also optimal if the environment becomes very hazardous (p close to 1). However, it turns out that a heterogeneous response with a (small) fraction of quickly responding cells is always optimal – no matter how dangerous the environment gets. Importantly, this result holds only if cells have $q \geq 2$ internal states, since this prevents cells from extinction after just one short sugar pulse.

Only in the case of one internal energy level, which corresponds to a very low sugar level in the environment, the optimal response switches from heterogeneous to homogeneous ($\alpha_{opt} = 0$) for dangerous environments with a high probability p of short sugar pulses. Mathematically,

Eq. (4.9) simplifies to

$$\begin{aligned}\lambda(\alpha) &= (1-p)\cdot\log_2\mu_{\mathrm{L}}(\alpha) \\ &\quad +(1-p)^2\cdot\sum_{n=1}^{\infty}p^n\log_2((1-\alpha)^n) \qquad (4.12)\\ &= (1-p)\cdot\log_2\mu_{\mathrm{L}}(\alpha)+p\cdot\log_2(1-\alpha)\,. \qquad (4.13)\end{aligned}$$

Therefore,

$$\frac{\partial\lambda}{\partial\alpha}\stackrel{!}{=}0=(1-p)\cdot\frac{\frac{\gamma}{\delta}-1}{\alpha\frac{\gamma}{\delta}+(1-\alpha)}-p\frac{1}{1-\alpha} \qquad (4.14)$$

can be solved analytically for α. We find $\alpha_{opt}>0$ for $p<p_{crit}=1-\frac{\delta}{\gamma}$ and $\alpha_{opt}=0$ for $p>p_{crit}$. Hence, the phase diagram for 1 internal state splits into a region where homogeneous timing is best and a region where heterogeneous timing is favorable (cf. Fig. 4.7C).

For this special case of only one internal state our model is very similar to an old problem in information theory, referred to as the Kelly problem of optimal gambling [86]. In this scenario an investor has a certain capital and decides to gamble on his stock repeatedly. In every round of the game he gambles with a fraction α of his capital and keeps the remainder $(1-\alpha)$ in a risk-less security. In case of winning a round with probability $(1-p)$ the invested fraction of his capital doubles, whereas it is lost with probability p. This model is analogous to our model with $\gamma/\delta=2$ and similarly predicts an optimal long-term growth rate at intermediate values of α. In contrast, Kelly also noted that after a finite number of rounds the expectation value $\langle N\rangle$ is maximized if the gambler bets all his money in every round ($\alpha=1$). While this strategy in fact optimizes short term growth, the probability of a single gambler to loose his entire capital (*gamblers ruin*) increases dramatically. Thus, the mixed strategy with $0<\alpha_{opt}<1$ optimizes the trade-off between short-term growth and long-term survival.

4.4.3. Discussion

Significance of Heterochronic Gene Induction

It is well established that snapshots of arabinose [161], lactose [130, 134] and galactose [2, 15, 85] operon expression display bimodal population responses at a given time after sugar addition. While this bimodality has first been reported in the lactose operon more than half a century ago, its physiological function for a bacterial community remained largely elusive. In contrast to, e.g., systems for immune evasion in pathogens, in which phenotypic heterogeneity is intimately related to the survival of the genotype [29], the evolutionary forces favoring heterogeneity in metabolic systems are not so obvious. For instance, in an environment of constant lactose abundance there exists an optimal Lac expression level that maximizes cell growth [39, 82], and fluctuations around this optimal level always reduce the growth rate [171]. While these models make predictions for enzyme expression in balanced growth, that is, after several generations of exponential growth when the macromolecular composition of a cell (e.g., the mass fractions of protein, RNA, and DNA) becomes stationary [158], in fluctuating environments growth is typically not balanced.

Under such conditions it has been suggested that a stochastic commitment of some fraction of the population to activate a metabolic network might provide selective advantages by anticipating the arrival of new food sources [146]. To our knowledge, however, wildtype carbon utilization systems are not activated spontaneously, i.e., in the absence of their natural

inducers, indicating that bacteria do not follow a 'pre-emptive' diversification strategy in their metabolic apparatus. Instead, cells directly sense the sugar in the environment and adjust the cell-to-cell variability in their response time to the detected sugar concentration. Our theoretical analysis revealed, that in a famine and feast scenario with stochastic durations of sugar availability, such responsive stochastic switching can maximize the long-term growth rate of a population by optimizing a trade-off between short-term growth and long-term survival. Hence, our results suggest that the observed heterochronic gene induction might serve as a regulated bet-hedging strategy in uncertain environments.

However, the model analyzed here considered only two different response times in an environment with two different sugar pulse durations. While this simplistic model allowed for an (at least partially) analytical treatment of the problem and also permitted some general conclusions about the functional advantages of a mixed strategy in fluctuating environments, it is (by construction) unable to predict the experimentally observed relation between the external sugar concentration and the width of the delay time distribution. Therefore, it will be interesting to test whether a generalized growth model with a continuous spectrum of delay times displays a similar scaling of population heterogeneity with the external sugar concentration turns out as the optimal regulation strategy. In fact, if the cell-to-cell variation of the delay time is smaller than the amortization period, it is possible that all cells of a colony suffer synchronously from a fatal series of sugar pulses, leading almost certainly to extinction in the long run. Hence, one might suspect that the width of the natural response-time distribution must be at least as broad as the amortization period in order to guarantee long-term survival of the population.

Gene Induction Under Energy-Limiting Conditions

One key ingredient to our model is that cells experience an amortization period after stimulation with low sugar concentrations, during which cells transiently invest more energy than they harvest by sugar digestion. While our estimate of the costs and benefits support this intuitive assumption, to date no experimental test has been performed on the fitness effects of transient sugar pulses. However, there are some indications suggesting that energy can become growth limiting under certain conditions. For instance, the value of the adenylate energy charge, which is proportional to the mole fraction of ATP plus half the mole fraction of ADP, is homeostatically controlled during exponential growth, but declines slowly during stationary phase in a wide variety of microorganisms [32]. Notably, during the slow decline in energy charge, all the cells are capable of forming colonies, whereas after prolonged starvation a steep drop in energy charge coincides with a rapid fall in viability [32]. These observations suggest, that energy is indeed a limiting factor for survival during stationary phase and that a careful usage of precious energy equivalents is essential for the ability to resume growth when nutrients suddenly become available.

In fact, for both prokaryotes and eukaryotes it has been observed that the addition of excess amounts of substrate to a previously starved population can cause a significant decrease in cellular viability [25, 45, 142]. This substrate-accelerated cell death has been related to an inherent risk of metabolic pathways with so-called 'turbo-design' [177]: Many catabolic pathways, such as glycolysis, begin with an ATP-requiring activation step, after which further metabolism yields a surplus of ATP. In the absence of specific regulatory mechanisms that prevent these pathways from demanding their toll before proper adaptation to the new environment has occurred, most available phosphate will be incorporated into glycolytic inter-

mediates [69], and the cell's ATP level gets too low to sustain cellular processes. Therefore, in order to prevent these hazardous situations, glycolysis has evolved a negative feedback on the initial activation step [180], thus turning glycolysis into a tremendously successful energy-generating strategy.

4.5. Conclusion and Outlook

In this chapter we characterized the molecular mechanisms of all-or-none gene expression in the arabinose utilization system by analyzing single cell fluorescence trajectories upon arabinose addition and removal. Our analysis revealed substantial delay and variation in the onset of gene induction, which seems closely related to the stochastic distribution of sugar uptake proteins at the time of inducer addition. In contrast, the down-regulation process upon arabinose removal was quick and homogeneous, and relies on the previously uncharacterized membrane protein AraJ. However, from our analysis we cannot exclude that there exists significant heterogeneity in AraJ expression, since switching from the on- to the off-state was much faster than the intrinsic timescales of GFP reporter expression and maturation. Potentially, also heterogeneous AraJ expression could in part account for heterogeneous timing in gene induction, since – as shown by our model – the delay time until full feedback activation depends mainly on the internal arabinose level, which is given by the ratio of arabinose import and export rates. Also, it is currently unclear how *araJ* is expressed in the presence and absence of arabinose and its chemical analogues, and further characterization of the *araJ* promoter will help to better understand the physiological role of sugar efflux in the arabinose utilization system.

Interestingly, similar temporal heterogeneity has been observed during the differentiation of *B. subtilis* [33, 38], where only a subpopulation of cells commits itself to sporulation [182]. Here the master regulator for sporulation is Spo0A, which is activated by phosphorylation via a phosphorelay cascade. Analogous to the heterogeneous expression of sugar transporters in the arabinose system, transcription of several phosphorelay genes is highly dynamic and variable in time, resulting in significant cell-to-cell variability of the phosphorelay phosphate flow and thus in a tremendous spread in the timing of sporulation initiation [33, 38]. This ensures that population heterogeneity is not generated instantly, but rather increases gradually in time and only if adverse environmental conditions persist [33]. Given that the response of many regulatory systems relies on sensors and signaling molecules available only in low copy numbers [12, 172], we expect that heterogeneous timing of cellular responses will be ubiquitously encountered at the single cell level.

While our theoretical analysis of the carbon utilization systems showed that temporal heterogeneity in gene induction can serve as a regulated bet-hedging strategy under certain conditions, further experimental verification of key model assumptions is clearly required. For instance, to test for the protein synthetic capacity of stationary phase cells, it will be crucial to probe a previously starved population for its induction behavior of the carbon utilization systems – especially at low sugar concentrations. Also, the hypothesis of a limited energy reservoir could be tested by measuring the viability of starved cells after exposure to an ever-changing sequence of short sugar pulses, as suggested in our famine and feast scenario. Quantitative time-lapse fluorescence microscopy in combination with current developments in microfluidics will provide powerful tools to get insight into the functional roles of temporal heterogeneity.

A. Publications

During the course of my PhD work several papers were published, four of which are subject of this thesis (Papers I–IV). The full texts of these articles are attached below.

Paper I G. Fritz, C. Koller, K. Burdack, L. Tetsch, I. Haneburger, K. Jung, and U. Gerland, 'Induction Kinetics of a Conditional pH Stress Response System in *Escherichia coli*', *J. Mol. Biol.*, **393**, 272 (2009)

Paper II I. Haneburger*, G. Fritz*, N. Jurkschat, L. Tetsch, A. Eichinger, A. Skerra, U. Gerland and K. Jung, 'The feedback-inhibitor cadaverine suppresses pH response by binding to the pH susceptible site of CadC in *Escherichia coli*', *submitted*

Paper III J. A. Megerle*, G. Fritz*, U. Gerland, K. Jung, and J. O. Rädler, 'iming and dynamics of single cell gene expression in the arabinose utilization system', *Biophys. J.*, **95**, 2103 (2008)

Paper IV J. A. Megerle*, G. Fritz*, S. A. Westermayer, D. Brick, R. Heermann, K. Jung, J. O. Rädler, and U. Gerland, 'Quantitative characterization of single cell switching dynamics in the arabinose utilization system', *submitted*

Paper V G. Fritz, N.E. Buchler, T. Hwa, and U. Gerland, 'Designing sequential transcription logic: a simple genetic circuit for conditional memory', *Systems and Synthetic Biology*, **1**, 89-98 (2007)

Paper VI K. Fritz, G. Fritz, B. Windschiegl, C. Steinem, and B. Nickel, 'Arrangement of Annexin A2 tetramer and its impact on the structure and diffusivity of supported lipid bilayers', *Soft Matter*, **6**, 4084-4094 (2010)

*equal contribution

doi:10.1016/j.jmb.2009.08.037 J. Mol. Biol. (2009) 393, 272–286

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Induction Kinetics of a Conditional pH Stress Response System in *Escherichia coli*

Georg Fritz[1,2], Christiane Koller[3], Korinna Burdack[3], Larissa Tetsch[3], Ina Haneburger[3], Kirsten Jung[3] and Ulrich Gerland[1,2]*

[1]*Arnold Sommerfeld Center for Theoretical Physics, Ludwig-Maximilians-Universität München, Theresienstraße 37, 80333 München, Germany*

[2]*Center for NanoScience, Ludwig-Maximilians-Universität München, Theresienstraße 37, 80333 München, Germany*

[3]*Munich Center of Integrated Protein Science at the Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Großhaderner Straße 2-4, 82152 Martinsried, Germany*

Received 18 May 2009; received in revised form 6 August 2009; accepted 13 August 2009
Available online 21 August 2009

Edited by M. Gottesman

The analysis of stress response systems in microorganisms can reveal molecular strategies for regulatory control and adaptation. In this study, we focused on the Cad module, a subsystem of *Escherichia coli's* response to acidic stress that is conditionally activated at low pH only when lysine is available. When expressed, the Cad system counteracts the elevated H^+ concentration by converting lysine to cadaverine under the consumption of H^+ and exporting cadaverine in exchange for external lysine. Surprisingly, the *cad* operon displays a transient response, even when the conditions for its induction persist. To quantitatively characterize the regulation of the Cad module, we experimentally recorded and theoretically modeled the dynamics of important system variables. We established a quantitative model that adequately describes and predicts the transient expression behavior for various initial conditions. Our quantitative analysis of the Cad system supports negative feedback by external cadaverine as the origin of the transient response. Furthermore, the analysis puts causal constraints on the precise mechanism of signal transduction via the regulatory protein CadC.

Keywords: CadB; CadA; lysine decarboxylase; acid stress response; quantitative modeling

Introduction

During their natural life cycle, gastrointestinal bacteria are faced with acid stress while passing the extreme low pH of the stomach and being exposed to volatile fatty acids in the intestine. *Escherichia coli*'s remarkable ability to sustain growth over multiple decades of H^+ concentrations[1] and its potential to survive extremely low pH are implemented by a battery of pH homeostasis[2–4] and acid tolerance[5–8] systems. In recent years, it was increasingly recognized that each of these subsystems is specifically activated under certain environmental conditions,[6,8] while the orchestration of the different responses is just beginning to be explored. However, a system-level study of the acid stress response requires detailed quantitative analysis of the individual modules.

One of the conditional stress response modules is the Cad system,[9–13] which is induced only when acidic stress occurs in a lysine-rich environment. The Cad module protects members of the *Enterobacteriaceae* against anorganic and organic acids in the intestinal tract[14] and against fermentation acids under phosphate-limiting conditions.[15] The three principal components of the Cad system are the enzyme CadA, the transport protein CadB, and the regulatory protein CadC (see Fig. 1). The decarboxylase CadA converts the amino acid lysine into cadaverine, a reaction that effectively consumes H^+.[16] The antiporter CadB imports the substrate, lysine, and exports the product, cadaverine. Together, CadA and CadB

*Corresponding author. E-mail address: gerland@lmu.de.

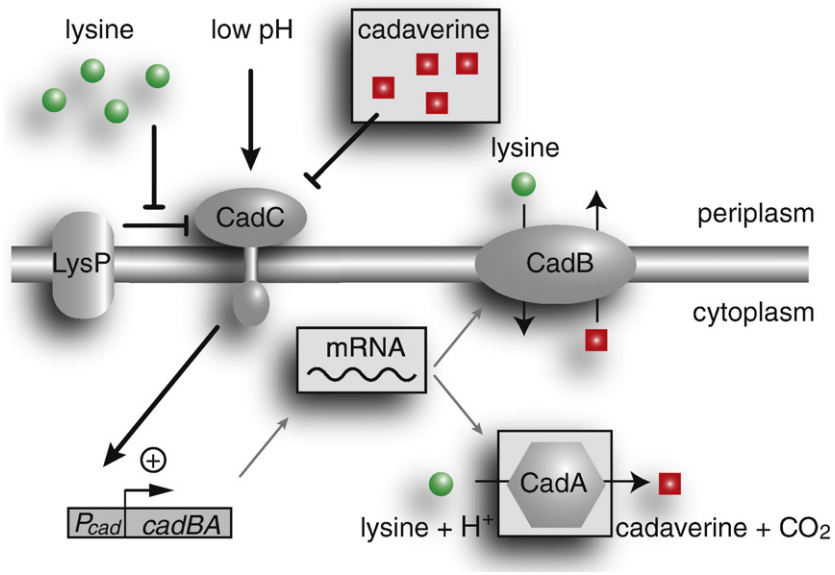


Fig. 1. Qualitative model of the Cad system in *E. coli* (simplified). The Cad system is conditionally activated by low pH and high lysine levels. Lysine inhibits the repressive effects of the lysine permease LysP on the receptor CadC, while low pH activates CadC directly. The active form of CadC activates transcription of *cadBA*, encoding the lysine decarboxylase CadA and the lysine/cadaverine antiporter CadB. The Cad system imports lysine, decarboxylates it under consumption of a cytoplasmic proton, and exports the product cadaverine in exchange for another lysine molecule. The net effects of these reactions are the expulsion of a proton from the cytoplasm and the excretion of the basic polyamine cadaverine. Finally, it is believed that external cadaverine deactivates CadC. In our experiments, we recorded the time evolution of the variables in gray boxes.

reduce the intracellular H^+ concentration and thereby contribute to pH homeostasis.[10,17] The cytoplasmic membrane protein CadC not only senses the external conditions[12,18,19] but also regulates the response by binding directly to the DNA and activating the transcription of *cadBA*.[20] Similar to other members of the ToxR family,[21] CadC thereby performs signal transduction in a single component, without the phosphorylation step employed by two-component systems.[22] Figure 1 also depicts the lysine permease LysP, which is not part of the *cad* operon but essential for its function since CadC senses lysine indirectly via interaction with LysP.[12,18,19,23] In contrast, the external (periplasmic) pH is believed to be sensed directly by CadC through a pH-dependent conformational transition and/or proteolytic cleavage.[19,24] The signal integration performed by CadC then ensures that CadA and CadB are produced only under the appropriate external conditions of low pH and lysine abundance.

However, CadC also senses a third input, which seems surprising from a physiological point of view: Dell *et al*. identified several CadC derivatives with single amino replacements that responded differently to cadaverine in comparison with wild-type CadC.[19] These data provided initial evidence that cadaverine influences CadC-mediated *cadBA* expression. Subsequently, it was found that cadaverine represses the long-term expression of the *cad* operon,[12] and we could show that the periplasmic domain of CadC in fact binds cadaverine *in vitro*.[23] As cadaverine is the end product of the decarboxylase reaction, it was suggested that it accumulates in the medium and causes a delayed transcriptional down-regulation of *cadBA* expression.[13] Although many stress response systems display a similar transient response,[25,26] their regulation strategy appears to be fundamentally different: for instance, the osmo-stress response of yeast directly follows its stimulus (low osmolarity) and remains active until the osmolarity returns back to physiological levels, while *cadBA* expression is down-regulated although the stress persists.

In this study, we explored whether negative feedback via external cadaverine can account for the transient response of the Cad system on a quantitative level. In particular, we were interested in the factors that determine the *duration* and the *amplitude* of the transient response and asked how the addition of external cadaverine affects these characteristic quantities. It is known, for instance, that external cadaverine reduces the long-term activity of the Cad system,[12] but one would like to know whether it shortens the duration of the transient expression pulse or reduces the amplitude of the pulse. Or does it affect both of these properties?

To address these questions, we quantitatively measured the dynamics of the Cad system in three important variables at high time resolution: the *cadBA* transcript, the activity of the lysine decarboxylase CadA, and the concentration of excreted cadaverine. Based on the existing qualitative model (cf., Fig. 1), we formulated a quantitative model for the Cad system and tested its agreement with the experimental response dynamics. We found that our quantitative model coherently describes the dynamical response of the wild-type Cad system within a physiological parameter regime. The available data constrain the key biochemical parameters to a narrow regime. For instance, we inferred the effective *in vivo* deactivation threshold for the Cad system and compared it with a previously measured *in vitro* binding threshold.[23] Using the quantitative model, we formulated predictions for the response dynamics of the Cad system under conditions with initially added cadaverine and in a mutant strain with a defective lysine permease, LysP. The suc-

cessful experimental validation of these predictions strongly supports the existence of the postulated feedback inhibition mechanism via cadaverine in the Cad system. Finally, we discuss the causal constraints of our results on the signal transduction mechanism by CadC, helping discriminate between two contradicting models.

Results

Transient expression dynamics

To probe the transient response of the Cad module, we first grew *E. coli* (strain MG1655) to exponential phase at pH 7.6 in minimal medium. We then induced the Cad module by transferring cells into fresh minimal medium with 10 mM lysine and buffered at pH 5.8. After shifting *E. coli* to inducing conditions, we monitored cell growth by determining the number of colony-forming units. Due to the acid stress, cell growth was significantly impaired, but cells remained viable over the 4-h period of the experiment (see Fig. S3). In the habitat of the intestines, fast and/or uniform growth of the cells is also not expected. Hence, the analysis of the Cad module under the present conditions is reasonable, as it mimics an *E. coli* population in its natural environment.

The induction defined the starting point, $t = 0$ min, for our measurements of the response, which we performed initially at intervals of 5 min until $t = 30$ min and then at longer intervals of 30 min. To quantify the response, we assayed the *cadBA* mRNA level, the specific CadA activity, and the external cadaverine concentration (see Materials and Methods for all experimental details). We found that the external pH, shown in Fig. 2a, remained low, even slightly decreasing from the induction level, over the entire 4-h period of the experiment. Transcription of the *cad* operon began immediately after induction, and mRNA rapidly accumulated, as shown in Fig. 2b. At $t \approx 25$ min, the mRNA level peaked and then rapidly decreased, reaching its low pre-induction level at about $t = 90$ min. The response on the protein level, as quantified by the specific activity of CadA shown in Fig. 2c, was slower, exhibiting a slight delay after induction and reaching a plateau level at $t \approx 25$ min, which was sustained over the time of the experiment. The activity of the Cad module led to the production and secretion of cadaverine, which accumulated in the medium, as shown in Fig. 2d.

The transient expression of the Cad module shown in Fig. 2 is in qualitative agreement with previous induction experiments that studied the system in a less quantitative manner.[13] The biochemical mechanism for the transient behavior remained unclear, however. It was suggested that the external cadaverine level exerts negative feedback on the activity of the regulator CadC.[12,13] Alternatively, the Cad module might, for instance, directly affect and control the level of its input stimuli. In other words, the activity of the Cad module might reduce the external lysine concentration below its induction threshold or shift the external pH level outside its range for induction.

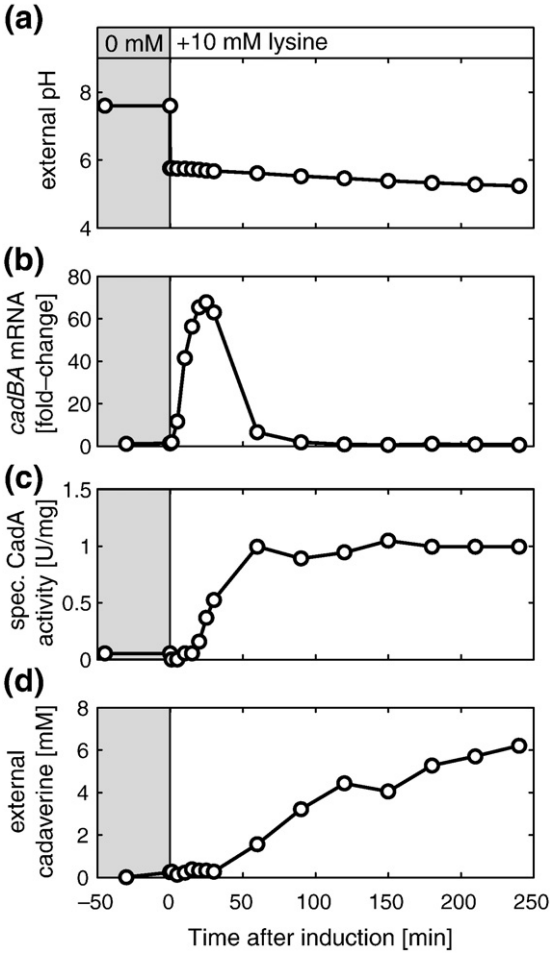


Fig. 2. Induction kinetics of the *cad* operon in *E. coli* MG1655. (a) The Cad system was induced at $t = 0$ min by a shift from pH 7.6 to pH 5.8 and simultaneous addition of 10 mM lysine. (b–d) The time evolution of the *cadBA* mRNA (b), the specific CadA activity (c), and the extracellular cadaverine concentration (d) was determined as described in Materials and Methods. All values are average values from duplicate repetitions. Specific CadA activity is given as U/(mg protein), with 1 U = 1 μmol cadaverine/min. For each sample, the cell density was adjusted to an optical density at 600 nm of 1, corresponding to 150 μg protein/mL. Concentration of total RNA was determined, and for each time point, equal amounts of RNA were tested with a radiolabeled probe directed against *cadBA* mRNA. Moreover, the signal intensity of *cadBA* transcript was compared with the signal intensity of *rpoD* mRNA that is constitutively transcribed. From these data, fold changes of *cadBA* transcription relative to the pre-induction value were calculated. For a detailed discussion, see the main text.

Dose–response curves

To address these possible alternative explanations and to characterize the ranges and the intensity of

the total response under our experimental conditions, we next determined the "dose-response" behavior of the Cad module. We have seen above (see Fig. 2c) that the CadA activity reaches a steady-state plateau about 60 min after induction. We took this plateau value as a proxy for the total (cumulative) response of the Cad system and studied its dependence on the input signals. To this end, we induced the Cad module with different external pH levels and initial lysine concentrations and assayed samples 90 min after induction for their CadA activity (see Materials and Methods for details). Panels (a) and (b) of Fig. 3 show the lysine dependence and the pH dependence of the response, respectively. The data in Fig. 3a indicate that when induced with pH 5.8, the Cad module is barely active at lysine concentrations below 0.5 mM, whereas it is fully active for lysine levels exceeding 5 mM. In between these values, the activity increases sigmoidally with the inducing lysine concentration. Similarly, at a given lysine induction level of 10 mM, the activity depends sigmoidally on the inducing pH (see Fig. 3b), with no significant activity above pH 6.8 and full activity at pH 5.8 and lower.

Taken together, the pH dependence of the total response shown in Fig. 3b and the time series in Fig. 2a show that the pH level did not leave the range for induction during the course of our experiment in Fig. 2. Hence, the transient behavior of the *cadBA* expression is clearly not mediated by a decrease of the external pH stimulus.

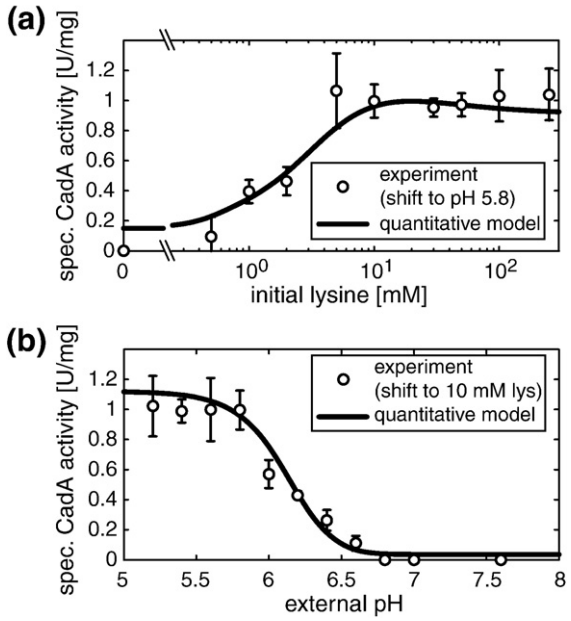


Fig. 3. Dose–response curves of the wild-type Cad system. (a) Dependence of the CadA response on the inducing lysine concentration. The specific CadA activity was determined 90 min after induction with pH 5.8 and the indicated lysine concentration (see Materials and Methods). Similarly, the dependence on the inducing pH in (b) was obtained by induction with 10 mM lysine and the indicated pH. Again, specific CadA activity was determined 90 min after induction. The continuous lines show the fit result of our quantitative model. Values are mean values from at least three independent experiments. Error bars represent standard deviation from these experiments.

The dose–response curves in Fig. 3 characterize the input–output behavior of the system when the Cad module is regarded as a "black box" signal processing unit. In particular, we can read off the apparent activation thresholds of the Cad module (i.e., the pH and lysine levels at which the module displays half-maximal activity). Such apparent thresholds constitute the first level of description in a top–down system analysis. Conversely, in a bottom–up analysis, the first level of description is via biochemical interaction parameters, while effective parameters, such as apparent thresholds, emerge from the interplay of molecular interactions. Quantitative modeling and analysis of this interplay are the only ways to make a connection between the two levels of description. In the following, we want to make such a connection and then leverage it to estimate the molecular activation thresholds of CadC from the apparent behavior of the Cad module.

Construction of a quantitative model

It is clear from the abovementioned data that a minimal quantitative model of the Cad module must describe the integration of the input signals pH and lysine, as well as the effect of cadaverine on the activity of CadC. Furthermore, it must describe the regulation and expression of the *cadBA* operon and the functioning of the CadA and CadB proteins, such that we may relate the quantitative model to the observed dynamics of the three system variables monitored in our experiments. Our construction of such a quantitative model is guided by the qualitative model in Fig. 1 and the known biochemistry of the Cad module. In that, our approach takes advantage of the wealth of prior knowledge about molecular interactions, which determine the topology (i.e., the "wiring") of the signaling network. In general, when the molecular players and their interactions are unknown, network reconstruction techniques can be used to gain qualitative and quantitative insights into the regulation of the system (see, for example, Refs. 27 and 28). However, here we are faced with a simpler problem where a predefined network topology has to be tested for agreement with the experimental data.

Signal integration

The three external signals known to affect the activity of the Cad module are the time-dependent lysine concentration, $l(t)$, the cadaverine concentration, $c(t)$, and the pH. The membrane protein CadC, which receives and combines these signals into a single response,[12,18,19] is constitutively expressed,[19] and hence we take the total amount of CadC per cell, C_0, to be constant. The signals then modulate only the fraction of active CadC molecules per cell, $C(t)/$

C_0. We assume that the different signals regulate CadC independently such that the CadC activity is described by the product form,

$$C(t)/C_0 = f(\mathrm{pH}(t)) \times g(l(t)) \times h(c(t)) \qquad (1)$$

Indeed, all experimental data available show no indication for a coupled effect of the input signals on CadC.[13,23] Another assumption implicitly made by Eq. (1) is that the fraction of active receptors is always equilibrated to the current levels of the input signals—that is, it does not depend on the signal levels prior to time t. This assumption is also plausible since the typical timescale for conformational transitions in receptors (see, for example, Ref. 29) is much shorter than the timescale of our experiments. The functions f, g, and h in Eq. (1) take on values between 0 and 1 and are assumed to be of the Hill form typical for cooperative binding reactions. The pH dependence is parameterized as follows:

$$f(\mathrm{pH}) = \frac{1}{1 + 10^{\frac{\mathrm{pH}-\mathrm{pH}_0}{\Delta\mathrm{pH}}}} \qquad (2)$$

with pH_0 denoting the pH value at which f reaches half-maximal activity and ΔpH determining the width of the response curve. Similarly, lysine dependence and cadaverine dependence respectively take the following forms:

$$g(l) = \frac{(l/K_l)^{n_l}}{1 + (l/K_l)^{n_l}} \quad h(c) = \frac{1}{1 + (c/K_c)^{n_c}} \qquad (3)$$

where K_l and K_c are the effective *in vivo* activation thresholds for the direct and indirect regulatory interactions of lysine and cadaverine with CadC, respectively. As usual, the Hill coefficients n_l and n_c parameterize the cooperativity of the binding reactions and determine the maximal sensitivity for signal detection. The difference in form between $g(l)$ and $h(c)$ stems from the fact that lysine activates the Cad module, whereas cadaverine represses it. On the other hand, the difference from Eq. (2) is due to pH being logarithmically related to the H^+ concentration.

Transcriptional regulation

In its activated conformation, CadC directly binds to the *cadBA* promoter P_{cad} and activates *cadBA* expression.[20] Generally, transcriptional regulation in bacteria can be described by quantitative "thermodynamic" models.[30,31] For the present case, an appropriate form for the transcriptional activity from P_{cad} as a function of the abundance of active CadC, $C(t)$, is derived in Materials and Methods. The resulting rate equation for the time evolution of the mRNA level m then takes the following form:

$$\frac{\mathrm{d}}{\mathrm{d}t}m(t) = \nu_\mathrm{m}\left(\frac{1 + (C(t)/K_C)^2 f}{1 + (C(t)/K_C)^2}\right)^2 - \lambda_\mathrm{m} m(t) \qquad (4)$$

with the basal transcription rate ν_m, the degradation rate λ_m, the fold change f between basal and maximal transcription rates, and K_C denoting the binding threshold for CadC-DNA binding. The particular choice of the exponents in the first term is motivated by the observation that the *cadBA* promoter appears to be regulated by two binding sites for dimeric forms of CadC.[20]

Kinetics of enzyme expression and catalysis

On the protein level, we have a similar interplay of synthesis and decay as in Eq. (4):

$$\frac{\mathrm{d}}{\mathrm{d}t}A(t) = \nu_\mathrm{p} m(t) - \lambda_\mathrm{p} A(t) \qquad (5)$$

where A is the abundance of CadA per cell and ν_p and λ_p are the translation rate and the degradation rate, respectively. The level of the transporter CadB is taken to be proportional to that of CadA since they are translated from the same mRNA. We neglect possible post-transcriptional regulation, for which there seems to be no experimental indication.[11] We also assume that we can subsume the transport and turnover of lysine to cadaverine through CadB and CadA by a single effective reaction since little is known about the microscopic rates and affinities of the coupled transport and decarboxylase reactions. As detailed in Materials and Methods, this assumption leads us to:

$$\frac{\mathrm{d}}{\mathrm{d}t}l(t) = -\nu_\mathrm{max} A(t) \frac{l(t)}{K_\mathrm{m} + l(t)} \qquad (6)$$

This simplified reaction corresponds to an effective Michaelis–Menten process with external lysine as the substrate, an effective maximal lysine turnover rate ν_max, and an effective Michaelis constant K_m. Also implicit in Eq. (6) is the assumption that growth of the bacterial population over the period of the experiment is negligible. For the external cadaverine level, we assume the flux balance,

$$\frac{\mathrm{d}}{\mathrm{d}t}c(t) = -\frac{\mathrm{d}}{\mathrm{d}t}l(t) \qquad (7)$$

implying that the sum of external lysine and cadaverine is conserved at all times, $l(t) + c(t) = \mathrm{const}$. This flux balance appears justified, given experimental results with a LysP-deficient mutant strain that we report and discuss further below. We take the pH to be a constant over the duration of our kinetic experiments since the pH changes only very little in our buffered medium and the Cad module is not very sensitive to the pH over this regime (see above).

Data interpretation with the quantitative model

We now demonstrate that the simple quantitative model constructed above is indeed a powerful tool. First, we tested the extent to which this model is compatible with the data sets reported above. For this test, we also included another data set from

Table 1. Repression of the long-term Cad response by cadaverine

Initial cadaverine	Relative *cadBA* expression	Model value
0 μM	1.00	0.97
20 μM	0.89	0.92
80 μM	0.60	0.74
320 μM	0.12	0.12
1300 μM	0.00	0.05

The data were taken from the work of Neely *et al.*[12] *cadBA* expression (central column), as determined from the β-galactosidase activity of a *cadA-lacZ* fusion, was measured in cells that were grown in medium, pH 5.8, with 10 mM lysine and the indicated cadaverine concentrations for 3 h.[12] The model values (rightmost column) show the fit result of our quantitative model.

Neely it *et al.*,[12] who determined the cadaverine-dependent dose response of the Cad module (see Table 1). This additional data set further constrains our model and probes consistency with the existing literature.

In total, the quantitative model has 14 parameters. We constrained each of these to a range inferred from typical physiological values and other information in the literature (see Table 2). We then fitted our model to all data sets simultaneously using standard least-squares minimization of the residual χ^2, as described in Materials and Methods. The curves corresponding to the best-fit parameters are shown in Figs. 3 and 4 (blue lines). The overall agreement with the experimental data is good, both for the response dynamics in Fig. 4a–c and for the dose–response curves in Fig. 3 and Table 1. Note, however, that not all model parameters are individually well constrained by the data. This becomes apparent by plotting the correlations between the quality of the fit, characterized by the residual χ^2, and the fit parameters, as shown in Fig. 5. As the fit becomes better (lower χ^2), most parameter values are confined to a narrow interval, indicating that the information contained in the experimental data accurately determines their values. For instance, the *cadBA* mRNA lifetime τ_m determines the decay time of the transient expression peak in Fig. 4a and is therefore strongly constrained in our model. In contrast, some of the parameters display a wide variation even at the lowest χ^2 values (e.g., the transcription and translation rates ν_m and ν_p). For these cases, where individual parameters are "sloppy,"[35] certain combinations of these parameters are well

Table 2. Parameters of the quantitative model

	Parameter	LB	UB	Estimated value	Comment
Sensory module					
K_l	Threshold for CadC activation by lysine	1	20	$3.6\pm^{5.8}_{0.6}$ mM	Bounds suggested by Fig. 3a
K_c	Threshold for CadC inactivation by cadaverine	50	1000	$235\pm^{32}_{49}$ μM	Bounds suggested by Fig. 3a
n_l	Hill exponent for CadC regulation by lysine	1	5	$1.1\pm^{0.2}_{0.1}$	
n_c	Hill exponent for CadC regulation by cadaverine	1	5	$2.8\pm^{0.9}_{0.3}$	
pH_0	pH threshold for CadC activation	—	—	6.2	Estimated from data in Fig. 3b
ΔpH	Width of the transition from active to inactive CadC	—	—	0.5	Estimated from data in Fig. 3b
Expression module					
C_0/K_C	Total CadC per cell in relation to the threshold for CadC-promoter binding	0.1	10	$1.1\pm^{2.6}_{0.1}$	The level of CadC is just sufficient to activate the pathway,[18] suggesting that *in vivo* $C_0/K_C \approx 1$
ν_m	Basal transcription rate	0.001	0.1	$4.3\pm^{14.1}_{2.2} \times 10^{-3}$ min^{-1}	
f	Fold change between basal and maximal transcription rates	10	1000	$698\pm^{170}_{452}$	Typical range[32,33]
ν_p	Effective translation rate	10^{-4}	10^{-1}	$4.2\pm^{8.6}_{2.3} \times 10^{-3}$ U/mg/min	Effective parameter with broad range
τ_m	mRNA half-life ($\ln 2/\lambda_m$)	1	50	$13.8\pm^{0.4}_{1.2}$ min	Typical range[34]
τ_p	Protein half-life ($\ln 2/\lambda_p$)	1	10^4	$29\pm^{2137}_{4}$ h	CadA is expected to be stable[16]
ν_{max}	Maximal rate for lysine turnover via CadA and CadB	10^{-4}	10	$1.3\pm^{1.4}_{0.5} \times 10^{-3}$ mM/min/(U/mg)	Effective parameter with broad range
K_m	Effective Michaelis constant for lysine turnover via CadA and CadB	1	100	$26\pm^{37}_{12}$ mM	Effective parameter with broad range

LB indicates lower bound; UB, upper bound. The estimated parameter values are shown as (best-fit value)$\pm^{\sigma^+}_{\sigma^-}$, where σ^+ and σ^- indicate the asymmetric standard errors in the positive direction and in the negative direction, respectively, see Eq. (14) in Materials and Methods.

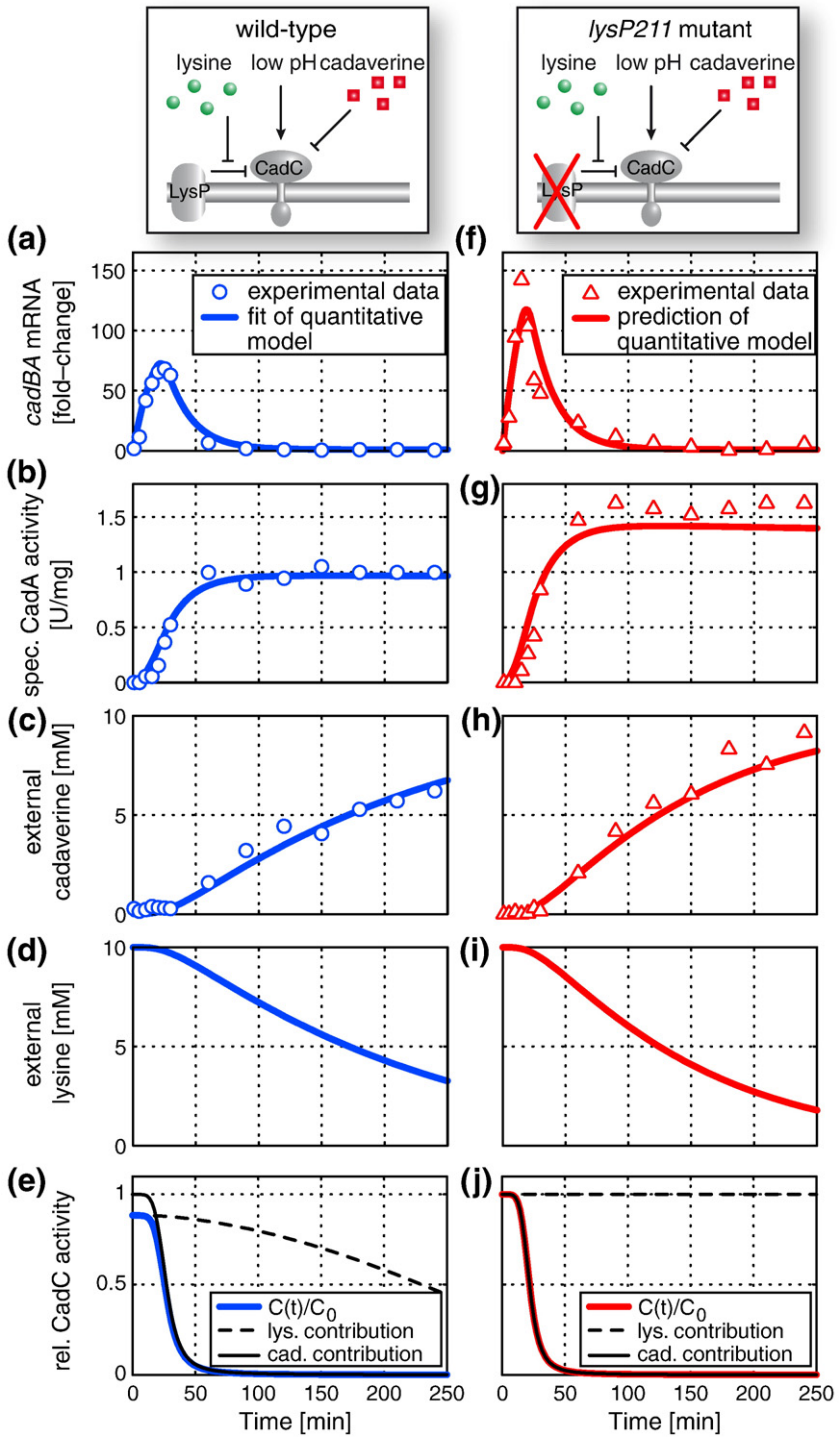


Fig. 4. Fit of our quantitative model (blue lines, left column) to the experimental data of the wild-type induction kinetics (blue circles) and the parameter-free prediction of our model for a LysP-deficient mutant (red lines, right column), together with the corresponding experimental data (red triangles), are shown. The external lysine concentration, shown in (d) and (i), and the relative CadC activity, shown in (e) and (j), have not been measured but were inferred from the quantitative model by a global fit to the wild-type data, as described in the main text and Materials and Methods. The dashed and continuous black lines in (e) and (j) show the lysine and cadaverine contributions, $g(l(t))$ and $h(c(t))$, in Eq. (3), to the signal integration function in Eq. (1).

constrained by the data sets. Pairwise scatter plots, as shown in Fig. S1, identify correlations and anti-correlations between the parameters and help to reveal the appropriate combinations. For instance, the product of the transcription and translation rates is much better determined by the data than the individual rates.

Please refer to Supplementary Material for the full discussion of our estimated parameter values in the light of previously published literature values. It is noteworthy that the best-fit value for the effective Hill coefficient n_c for the regulation of CadC by cadaverine is close to 3 and is relatively well constrained by the data. This suggests that a molecular mechanism for cooperativity is at work, possibly a multimerization of CadC proteins in the membrane. Another interesting observation from Table 2 concerns the half-life of *cadBA* mRNA, which was well constrained by the data to a value of almost 14 min. A global analysis of RNA half-lives in *E. coli*[36] found an extremely short half-life of less than 2 min for the *cadBA* mRNA, suggesting that an active degradation mechanism is involved. Our Northern blot data for the *lysP211* mutant, shown in Fig. 4f, do indeed

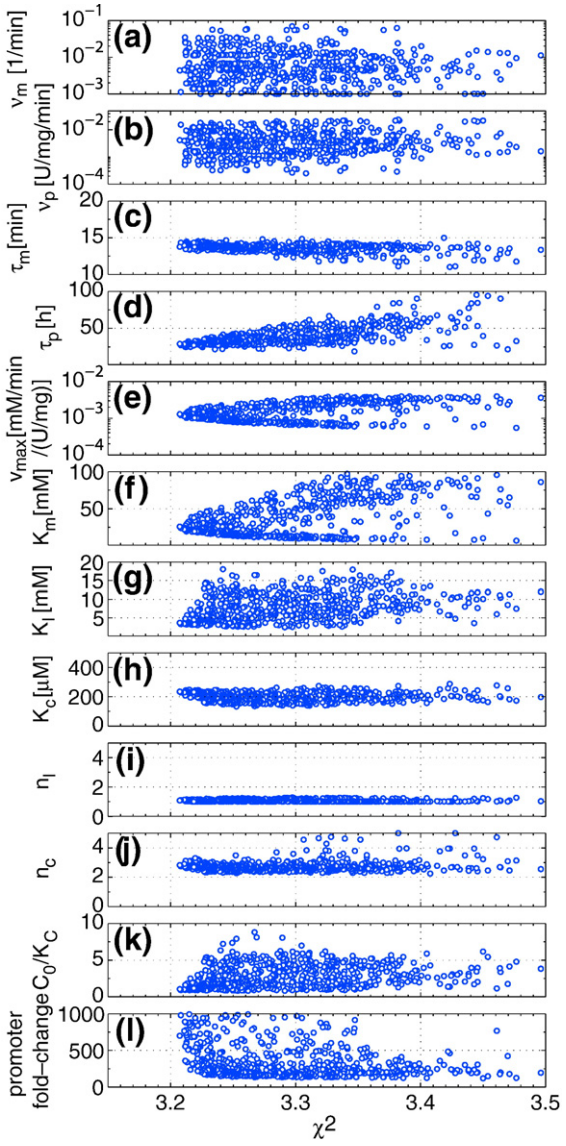


Fig. 5. Correlations between the goodness of fit and the estimated parameters. The points correspond to local optima in the parameter space, for which the difference between the quantitative model and the experimental data in Fig. 2, Fig. 3, and Table 1 is minimized (see Materials and Methods for all details). As the fit quality increases (lower χ^2), most parameters are confined to narrow intervals, indicating that their values are well constrained by the experimental data. However, some parameters display significant variation even for the lowest χ^2 values and one finds, from parameter–parameter correlation analysis in Fig. S1, that only combinations of those are well confined by our data.

suggest a rapid decay of the mRNA at high levels directly after the peak, followed by a slower decay at lower levels. Our quantitative model only allows for a single degradation rate, which leads to the intermediate half-life of 14 min as a best-fit value. The changing degradation rate could be rationalized under the assumption that the *cadBA* mRNA has a relatively weak binding affinity to the degrading enzyme such that active degradation only contributes significantly at high mRNA levels, whereas a slower passive decay is at work at low mRNA levels. Figure S2 illustrates that this mechanism could indeed account for the observed shape of the mRNA curve in Fig. 4f. However, this explanation would raise the question of why the rapid active decay is not observed in the data of the wild type in Fig. 4a as such. Possibly, the kinetics of LysP unbinding from CadC is slow (contrary to our model assumption of a rapid binding equilibrium) such that it interferes with the reception of the cadaverine signal and thereby broadens the onset of transcriptional down-regulation. Further experiments would be required to test these hypotheses.

Given the compatibility of the model and data, we next used the model to infer quantitative characteristics of the module that are not directly assayed. For instance, the activity of the central regulator CadC as a function of the external signals lysine, pH, and cadaverine is a biochemical characteristic that is pivotal to the function of the module but difficult to measure directly. In our model, this quantitative characteristic is represented by the signal integration function, Eq. (1). On the other hand, the experimental data shown in Fig. 3 probe the final output of the Cad module on the protein level (specific CadA activity), which is the integrated result of a nonlinear dynamical system with feedback. Figure 6 compares this final system output (black curves, as in Fig. 3, but normalized to 1) with the inferred activity of the CadC regulator (blue curves). The latter represents the three sigmoidal functions that make up the signal integration function, Eq. (1), with the parameters determined from the global fitting procedure described above. We observe from Fig. 6 that the final system output behaves qualitatively similar to the inferred biochemical activity of CadC. However, in each case, the apparent activation threshold for the system response (point of half-maximal CadA activity, black curve) is shifted with respect to the inferred biochemical activation threshold (point of half-maximal CadC activity, blue curve). These shifts are due to the fact that the total CadA activity depends not only on the characteristics of the regulator but also on the biochemical properties and timescales of the negative feedback loop. In principle, the feedback can even lead to nonmonotonic behavior in the dose response, despite the underlying monotonic dependence of the signal integration function on the levels of the external signals†. The inferred *in vivo* biochemical activation thresholds, $K_I = 3.6$ mM and $K_c = 235$ μM, and the values of the Hill coefficients describing the sensitivities to the signals can be read off directly from the blue curves, as indicated in Fig. 6.

It is useful to compare these biochemical activation thresholds with the actual concentrations

† The model predicts indeed a weak nonmonotonic effect in the lysine dependence; however, this is not a robust prediction.

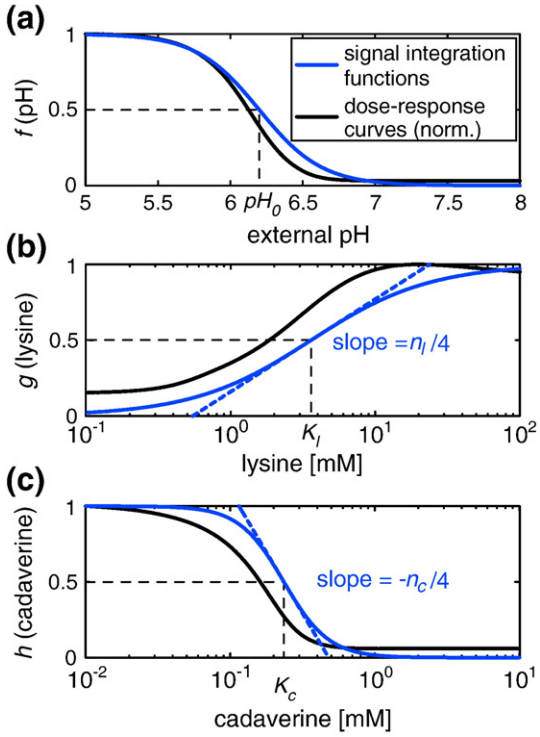


Fig. 6. Comparison of the extracted signal integration functions (blue curves) with the normalized dose–response curves (black curves). The signal integration functions describe the dependence of CadC activity on pH level (a), on lysine (b), and on cadaverine (c). They correspond to functions f, g, and h in Eqs. (2) and (3) and are plotted for the best-fit parameters listed in Table 2. The apparent activation thresholds of the dose–response curves (black curves) are shifted with respect to the inferred biochemical thresholds of CadC (blue curves) since the dose–response curves also depend on the biochemical properties of the nonlinear feedback in the Cad module. From the extracted signal integration functions, we can also read off the sensitivity of CadC on its input signals since their maximal slopes are determined by the Hill coefficients n_l and n_c.

encountered in our kinetic induction experiments, in particular at the time of transcriptional down-regulation ($t = 30$ min). From the plots in Fig. 4c and d, we see that the lysine concentration at this time point is still about a factor of 3 higher than the activation threshold, whereas the cadaverine concentration of 300 μM exceeds the deactivation threshold. The individual regulatory contributions from lysine and cadaverine to the CadC activity [i.e., $g(l(t))$ and $h(c(t))$] are plotted in Fig. 4e. The lysine curve (dashed line) displays only a very weak impact on the CadC activity, whereas the increase of cadaverine is the primary effect causing the down-regulation of the CadC activity (continuous line). Hence, the analysis with our quantitative model strongly suggests that the negative feedback via external cadaverine can quantitatively explain the timing of the transient response in the wild-type Cad system and that the decreasing lysine stimulus is not involved in this behavior.

Prediction and experimental analysis under altered conditions

We have so far analyzed the Cad module only in the wild-type strain and only with a single induction protocol. To obtain a more complete picture of its quantitative behavior, we constructed a mutant strain, MG1655-*lysP211*, with a truncated and inactive form of the lysine permease LysP (see Materials and Methods for details). Qualitatively, we expected that this mutation would completely abolish the lysine requirement for the activation of CadC since the truncated form of LysP would be unable to repress CadC[23] (an early study also indicated a derepressed activation of the Cad system by a spontaneous mutation in *lysP*[9]). Within our quantitative model, the *lysP211* mutation was mimicked by setting the lysine-dependent activity function in Eq. (1) equal to its maximal value [i.e., $g(l) = 1$], thereby rendering it independent of the inducing lysine concentration. Apart from this "*in silico* mutation," we left the model and the parameter values unchanged.

We verified that the LysP-deficient mutant was indeed inducible by a shift from pH 7.6 to pH 5.8 alone and did not require lysine for its induction (data not shown). We then performed the same kinetic induction experiments as with the wild-type strain [see the results in Fig. 4f–j for the *lysP211* mutant (red triangles) and the quantitative model (red curves)]. We clearly saw that the expression in the mutant remains transient, again supporting the conclusion that a signal different from lysine mediates the transcriptional down-regulation. The shape of the response is altered, however, with a stronger initial expression and a peak in the mRNA level that has a larger amplitude and reaches its maximum earlier than for the wild-type strain. Exactly these features are expected also on the basis of the quantitative model: The stronger initial expression is due to the full relief of the LysP repression, and the negative feedback via cadaverine sets in at an earlier time since the cadaverine threshold is more rapidly reached (see Fig. 4h and j). Also, the CadA activity, shown in Fig. 4g, is expected to reach a higher steady-state plateau, as observed in the experiment. Interestingly, at the end of the experiment, the cadaverine level in Fig. 4h reaches almost the 10 mM level of initially added lysine, which is in line with our flux balance assumption in the model (see previous section).

Next, we considered altered environmental conditions for the induction of the Cad module. Neely *et al.*[12] had already shown that induction with initially added cadaverine causes a significantly reduced long-term activity of the *cad* operon (cf., Table 1). Yet, it is not clear whether this diminished long-term activity is caused by a *cadBA* expression pulse of similar strength but reduced duration, by a pulse of reduced strength with similar duration, or by a combination of both. To resolve this question, we again performed kinetic induction experiments with the wild-type strain under identical conditions

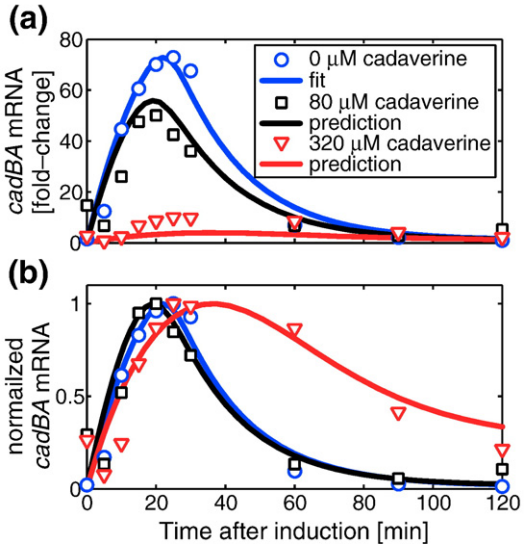


Fig. 7. Experimental test of the kinetic model prediction. The lines show the parameter-free model prediction for induction of the Cad system by a shift to pH 5.8 and 10 mM lysine, together with the cadaverine concentration indicated in the legend. The experimental data in (a) (symbols) were recorded and scaled as described in Materials and Methods. In (b), all data were normalized to their maximal values.

as in Fig. 2 but additionally with 80 or 320 μM cadaverine supplied at the time of induction. In the latter case, the initially supplied cadaverine already slightly exceeded the inferred inactivation threshold of K_c=235 μM such that a strong effect on the response could be expected. The resulting data are shown in Fig. 7a (squares and triangles), together with the original data (no added cadaverine, circles) for comparison. Figure 7b shows the same data but with all curves normalized to peak height 1 in order to emphasize the shape of the response. We observed that the primary effect of the addition of initial cadaverine is to reduce the strength of the response. This is also predicted by the quantitative model (continuous lines), rather accurately for the 80 μM cadaverine data set, while the reduction for 320 μM cadaverine is predicted to be stronger than that observed experimentally.

It should be noted that transcription was detected by Northern blot analysis, so that the shape of the response could be more accurately determined than the absolute amplitude. It is then interesting to observe that the quantitative model predicts a widening of the expression peak and a shift of the maximum to a later time for the highest cadaverine concentration (320 μM, red curve). Again, this behavior results from the nonlinear feedback in the Cad module—that is, the addition of initial cadaverine decreases the initial rate of mRNA production such that the CadA level increases more slowly and the negative feedback also sets in more slowly. This predicted change in the shape of the expression peak for 320 μM agrees remarkably well with the experimental observation (see Fig. 7b). This finding provides strong evidence that the quantitative characteristics of the Cad module are well described and understood with the help of our mathematical model.

Discussion

Conditional stress response with feedback inhibition

In this work, we analyzed the kinetics of a conditional pH stress response system, the lysine decarboxylase system of *E. coli*, which exhibits only transient induction, even when pH stress persists. Our results strongly suggest that the additional stimulus for the conditional response (i.e., a lysine-rich environment) is also not responsible for the transient behavior. Rather, our kinetic and dose-response experiments in combination with our quantitative model clearly indicate that negative feedback via the product of the decarboxylation reaction, cadaverine, leads to the down-regulation of the response.

Cadaverine has previously been linked to the transient behavior in pivotal work on this stress response system.[12,13] This link was based on observations that external addition of cadaverine significantly reduces the long-term Cad activity and that a CadA$^-$ mutant displayed persistent *cadBA* expression. However, if the transcriptional shutoff was mediated by a decrease of the lysine stimulus, the persistent *cadBA* expression could alternatively be explained by the lack of lysine consumption in the CadA$^-$ mutant. Also, while a reduction in the steady-state activity implies that the Cad system is generally repressed by cadaverine, it is not clear that the timing of its down-regulation is set via this negative regulatory interaction. None of the previous studies directly measured the system-induced dynamics of the cadaverine concentration or studied the kinetics of the Cad system with externally added cadaverine. By performing these quantitative experiments with a high time resolution and by interpreting them with a quantitative theoretical model, we obtained evidence for a causal relation between the time of transcriptional down-regulation and the increase of the external cadaverine concentration above its deactivation threshold. For instance, the decreasing amplitude in the dynamical response after adding external cadaverine and a more subtle delayed down-regulation, both predicted by the quantitative model, were strikingly confirmed by our kinetic measurements.

Signal transduction mechanism

Our system-level study of the Cad module also permits some conclusions about the involved molecular interactions and the signal transduction mechanism. Using our quantitative model, we were able to estimate the relevant *in vivo* activation and

deactivation thresholds of the regulatory protein CadC. For the inactivation of CadC by external cadaverine, we found a threshold of $K_c = 235$ μM. This value is surprisingly close to the *in vitro* binding constant of 96 μM for the interaction of cadaverine with the periplasmic domain of CadC.[23] In contrast, CadC has almost no affinity for lysine. There is recent evidence that CadC is inhibited at low lysine concentrations via a transmembrane domain interaction with the lysine permease LysP, whereas the interaction is released at high lysine levels.[23] In the present work, we determined the effective *in vivo* lysine activation threshold to be $K_l = 3.6$ mM. This result is somewhat surprising since the Michaelis constant K_m for lysine transport by LysP is much lower at ~10 μM.[37] However, the LysP–CadC interaction and the K_m of LysP do not necessarily need to have a direct correspondence.

An interesting open question concerns the signal transduction mechanism of CadC. Two alternative models have been proposed: (i) a reversible conformational transition of CadC activates its cytoplasmic N-terminal domain and allows it to bind to the promoter while remaining integrated in the membrane and (ii) a shift to acidic pH induces cleavage of the cytoplasmic domain, allowing it to diffuse freely to the promoter. The existence of the negative feedback by external cadaverine puts causal constraints on these microscopic mechanisms. If the cleavage mechanism were realized, it is not clear how external cadaverine could down-regulate *cadBA* expression after it has been induced since the freely diffusing cytoplasmic domains would no longer be able to recognize this signal. A high turnover of cleaved CadC could solve this problem by rapidly eliminating unresponsive activators. However, the increased degradation of CadC under inducing conditions would have to be balanced by an elevated *cadC* expression. This, however, is in disagreement with previous observations where it was found that *cadC* expression is constitutive.[18,19] Hence, the existence of the negative feedback by external cadaverine leads us to favor the reversible model.

Top–down system analysis of a functional module

The Cad system of *E. coli* is a functional module with few closely connected molecular components. How such modules integrate, process, and respond to external signals is a central question—but generally also a difficult one. The approach taken in the current study is akin to a "top–down" system analysis, where input signals are controlled and the output(s) and key internal system variables are measured. This is in contrast to a biochemical "bottom–up" approach, where each component would first be characterized separately, followed by their pairwise interactions, gradually moving upward in complexity. In our study, we were able to bring these two complementary approaches into a first contact for the Cad module, with the help of our quantitative model that provided the means to estimate relevant *in vivo* values for biochemical interaction parameters and the quantitative form of the signal integration function displayed in Fig. 6. Without the quantitative model, we would have been unable to extract this "hidden information" from the experimental data. Direct *in vivo* measurements of signal integration functions have been performed for the regulatory circuit controlling chemotaxis in *E. coli* using sophisticated single-molecule techniques. Since these powerful techniques are not easily transferred to the large class of functional modules of interest, the indirect approach taken in the present work may often be a welcome alternative.

Conclusions and outlook

Our quantitative analysis of the Cad system provides a first step toward a system-level understanding of the complex acid stress response network in microbes. Analogous quantitative studies of the two other major amino acid decarboxylase systems, glutamate and arginine decarboxylase,[8] could reveal important insights into how these modules are orchestrated in the complex environment of their host. The presence of multiple amino acids in the natural environment of *E. coli* suggests that these conditional stress response systems are often induced in parallel. It will be interesting to study how these systems are coordinated to provide an effective and robust way of pH homeostasis and acid tolerance response.

Materials and Methods

Bacterial strains and growth conditions

E. coli MG1655[38] was used as wild-type strain. The *lysP211* mutant was obtained via undirected mutagenesis.[9,39] For this purpose, cells of *E. coli* MG1655 were grown on minimal agar plates with 0.2% (w/v) glucose as sole carbon source[40] containing 100 μg/mL of thiosine (*S*-aminoethyl cysteine), a toxic lysine analog that leads to spontaneous mutations in *lysP*. One mutant, designated MG1655-*lysP211*, had a nucleotide exchange at position 211 in *lysP*, resulting in a stop codon and hence a truncated and inactive form of LysP (70 amino acids). *E. coli* strains MG1655 and MG1655-*lysP211* were grown aerobically in shaking flasks under noninducing conditions at pH 7.6 in 5 L of phosphate-buffered minimal medium[41] containing 0.4% (w/v) glucose as sole carbon source at 37 °C to an optical density at 600 nm of 0.5 (noninducing conditions). Subsequently, cells were collected by centrifugation (10 min, 4000*g* at 37 °C) and transferred into fresh prewarmed minimal medium, pH 5.8, containing 10 mM L-lysine (L-lysine-hydrochloride, Roth) and 0.4% (w/v) glucose in a 5-L fermenter (Biostat B, Satorius BBI Systems GmbH) (inducing conditions). Cultivation of cells was continued anaerobically at 37 °C. At the indicated times, samples were taken and then centrifuged at 4000*g* (4 °C) for 5 min, and cell pellets as well as supernatants were separately stored at −80 °C until further use. The number of colony-forming units was determined after incubation of 100 μL of various dilutions on LB agar plates overnight at 37 °C.[42]

Lysine decarboxylase assay

Specific activity of the lysine decarboxylase CadA was measured by resuspending cells corresponding to 10 mL of culture in 1 mL of Ldc buffer [100 mM Na-acetate, pH 6.0, 1 mM ethylenediaminetetraacetic acid, 0.1 mM pyridoxal phosphate, 10 mM β-mercaptoethanol, 10% (w/v) glycerol]. Lysozyme (0.1 mg/mL) was added, and the mixture was incubated on ice for 30 min. The cell suspension was sonified (10%–20% amplitude; 0.5-s output pulse; digital Branson Sonifier II 250), and the lysate was centrifuged at 15,000*g* at 4 °C for 15 min. Activity of lysine decarboxylase in cell-free extracts was measured as described previously[43] using 5 μg of protein per assay. Specific activity is defined as 1 U/mg = 1 μmol cadaverine/(min × mg protein).

Measurement of extracellular cadaverine

The extracellular cadaverine concentration was determined according to the spectrophotometric method described by Phan *et al*.[44] Briefly, 10 μL of culture supernatant was diluted 5-fold with H_2O_{dest}, 120 μL of Na_2CO_3 (1 M) and 120 μL of TNBS (2,4,6-trinitrobenzene-sulfonic acid; 10 mM) were then added, and the mixture was incubated for 4 min at 40 °C. After extraction with 1 mL of toluene, the absorption of the organic phase (containing *N*,*N*-bistrinitrophenylcadaverine) at 340 nm was measured. The cadaverine concentration was calculated based on a standard curve using cadaverine dihydrochloride (Sigma) between 0 and 500 nmol.

Preparation of RNA

Total RNA was isolated according to the method of Aiba *et al*.[45] Briefly, cells were resuspended in cold 20 mM Tris–HCl, pH 8.0, and subsequently lysed by addition of 20 mM sodium acetate, pH 5.5, 0.5% (w/v) SDS, and 1 mM ethylenediaminetetraacetic acid, pH 8.0. Then, RNA was extracted with prewarmed (60 °C) acid phenol, and the mixture was centrifuged at 12,000*g*. After an additional extraction of RNA using phenol/chloroform/isoamyl alcohol (25:24:1), RNA was precipitated with 100% ethanol at −20 °C overnight. The precipitate was washed with 70% (v/v) ethanol, and the dry RNA pellet was dissolved in 35 μL of H_2O_{dest}. RNA concentration was determined by measuring the absorption at 260 nm. All solutions were prepared with 0.1% (v/v) DEPC (diethylpyrocarbonate).

Northern blot analysis

cadBA mRNA was quantitatively determined by Northern blot analysis. This technique does not allow determination of the absolute number of *cadBA* transcripts, but it determines the relative induction level (ratio between induced cells and pre-induced cells) by using a radioactive labeled probe that specifically detects *cadBA* mRNA. Nevertheless, the readout scales linearly with *cadBA* mRNA concentration. Northern blot analysis was performed following a protocol described earlier.[46] Briefly, 20 μg of RNA was separated by electrophoresis in 1.2% (w/v) agarose–1.1% (v/v) formaldehyde gels in Mops (4-morpholinepropanesulfonic acid) buffer. RNA was transferred to a Hybond nylon membrane (GE Healthcare) by capillary blotting. Hybridization was performed following a standard protocol[42] using an [α-^{32}P]dCTP-labeled PCR fragment of the first 400 base pairs of *cadBA*. Radioactive labeling was quantified with a PhosphoImager. As control, expression of *rpoD*, a housekeeping gene of *E. coli*, was analyzed. Signal intensity of *cadBA* mRNA was normalized to the signal intensity of *rpoD* mRNA. If not indicated otherwise, the data are given as fold change of *cadBA* transcription relative to the pre-induction value. Additionally, in Figs. 4 and 7, the absolute magnitude of the mRNA fold change was re-scaled such that the integral over the expression curve was proportional to the long-term CadA activity.

Model details

From thermodynamic models of transcriptional regulation reviewed by Bintu *et al*.,[30,31] the effective transcription rate, ν_m^{eff}, as a function of two activators, *A* and *B*, with independent binding sites (binding constants K_A and K_B) is given by

$$\nu_m^{eff} = \nu_m \left(\frac{1 + (A/K_A) f_A}{1 + (A/K_A)}\right)\left(\frac{1 + (B/K_B) f_B}{1 + (B/K_B)}\right) \qquad (8)$$

cf., Ref. 30, Table 1, case 10. We make the simplifying assumption that the binding sites are identical and that

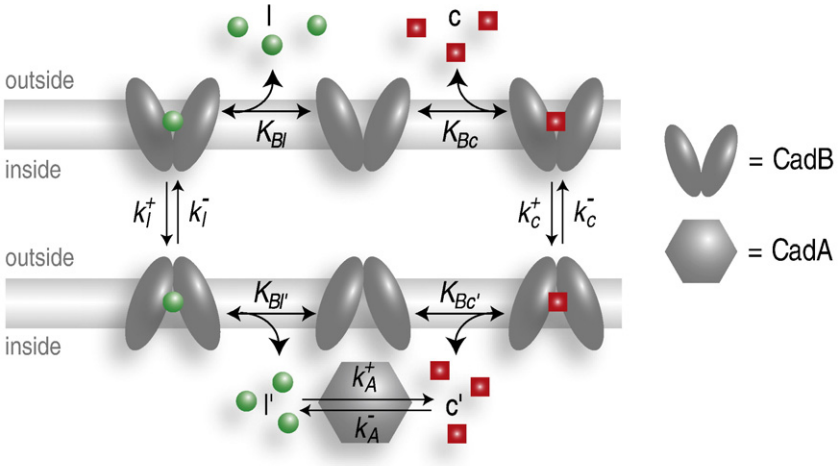


Fig. 8. Kinetic scheme of lysine and cadaverine transport and turnover. Antiport of lysine (*l*) and cadaverine (*c*) by CadB is modeled in analogy to the homologous arginine–ornithine antiporter ArcD in *L. lactis* by a single-site Ping Pong Bi–Bi mechanism.[47] Interconversion of lysine and cadaverine by CadA is modeled by a reversible first-order reaction.[48] The prime (′) indicates internal quantities. The rates (k_l^+, k_l^-, k_c^+, k_c^-, k_A^+, k_A^-) and equilibrium constants (K_{Bl}, $K_{Bl'}$, K_{Bc}, $K_{Bc'}$) are indicated next to the reaction steps.

each site can only be bound by a dimer of CadC (C_2).[20] Setting $A=B=C_2$, exploiting mass action $K=C^2/C_2$, and introducing the effective binding constant $K_C=\sqrt{KK_{C_2}}$ lead to the first term in Eq. (4).

The effective Michaelis–Menten form of the lysine turnover rate in Eq. (6) was derived as follows: The transport of lysine and cadaverine via CadB was modeled in analogy to the homologous arginine–ornithine antiporter ArcD in *Lactococcus lactis*[47] (see Fig. 8). For low external cadaverine and internal lysine concentrations, a general form of the inwardly directed flux of lysine is as follows:[47]

$$v_{\text{eff}} = k_l^+ B \frac{[\text{l}]}{K_{Bl} + \left(1 + \frac{k_l^+}{k_c^-} + \frac{k_l^+ K_{Bc'}}{k_c^- [\text{c}']}\right)[\text{l}]} \tag{9}$$

where B is the number of CadB molecules per cell and [l] and [c] are the concentrations of lysine and cadaverine on the outer surface of the membrane, respectively. Internal solute concentrations are marked by a prime (′). The parameters k_l^+, k_c^-, K_{Bl}, and $K_{Bc'}$ are defined in the kinetic scheme of Fig. 8. Here, it is assumed that (i) the conformational transition in CadB, which mediates the transport, does not occur without bound lysine or cadaverine, (ii) the antiporter and its substrates are in binding equilibrium at each surface of the membrane, and (iii) the membrane translocation reaction of the carrier is slow and rate limiting. The internal cadaverine concentration [c′] is determined by an interplay of the (reversible) decarboxylation through CadA[48] and by the export via CadB. In steady state, we find

$$[\text{c}'] = \frac{k_A^+}{k_A^-}[\text{l}'] - \frac{v_{\text{eff}}}{k_A^- A} \tag{10}$$

where A is the number of CadA molecules per cell. If the equilibration between lysine and cadaverine through CadA is fast compared with the transport through CadB, the second term is negligible and the internal cadaverine level is solely determined by the internal lysine level [l′]. In steady state, it turns out that [l′] is not affected by the Cad module, since the 1:1 stoichiometry of the antiporter ensures that lysine decarboxylation is balanced by lysine import. If we further take advantage of the fact that CadB and CadA are transcribed polycystronically and that there seems to be no post-transcriptional regulation,[11] we can set CadB proportional to CadA (i.e., $B=\alpha A$). Taken together, in the limit of rapid CadA kinetics and low external cadaverine and internal lysine concentrations, Eq. (9) reduces to the simple effective Michaelis–Menten form in Eq. (6) with CadA as the enzyme, $v_{\max}=(\alpha\times k_l^+)/\eta$ as the effective maximal turnover rate, and $K=K_{Bl}/\eta$ as the effective Michaelis constant, where $\eta = 1 + \frac{k_l^+}{k_c^-} + \frac{k_l^+ k_A^- K_{Bc'}}{k_c^- k_A^+ [\text{l}']}$.

Parameter estimation

The parameters of our quantitative model were estimated by using a trust-region reflective Newton method (MATLAB, The MathWorks, Inc.) to minimize the total χ^2, defined by

$$\chi^2(\vec{\theta}) = \chi^2_{\text{kin}}(\vec{\theta}) + \chi^2_{\text{pH}}(\vec{\theta}) + \chi^2_{\text{lys}}(\vec{\theta}) + \chi^2_{\text{cad}}(\vec{\theta}) \tag{11}$$

with respect to the parameter vector $\vec{\theta} = (\theta_1, \ldots, \theta_M)$, where $M=14$ is the total number of model parameters (cf., Table 2). The contribution of the kinetic data is calculated from

$$\chi^2_{\text{kin}}(\vec{\theta}) = \sum_{i=1}^{3} \frac{1}{N_i \overline{\Delta t_i}} \sum_{j=1}^{N_i} \Delta t_{ij} \left(\frac{y_{ij} - \tilde{y}_{ij}(\vec{\theta})}{\sigma_{ij}} \right)^2 \tag{12}$$

where y_{1j}, y_{2j}, and y_{3j} are the experimental data of the *cadBA* mRNA, the CadA activity, and the cadaverine concentration at time t_j, respectively. Similarly, $\tilde{y}_{ij}(\vec{\theta})$ denotes the corresponding values of the quantitative model for a given parameter set $\vec{\theta}$ and σ_{ij} is the standard error of each measurement (estimates from our experiments: $\sigma_{1j}=5$, $\sigma_{2j}=0.1$ U/(mg protein), and $\sigma_{3j}=0.5$ mM $\forall j$). N_i is the number of data points of a given time series, $\Delta t_{ij} \equiv (t_{i(j+1)} - t_{i(j-1)})/2$ is the time between subsequent data points, and $\overline{\Delta t_i}$ is the mean time between the data points in data set i. The contributions of the dose–response curves are similarly defined and are exemplarily shown for the pH-dependent response:

$$\chi^2_{\text{pH}}(\vec{\theta}) = \frac{1}{N_{\text{pH}}} \sum_{i=1}^{N_{\text{pH}}} \left(\frac{A_\tau(\text{pH}_i) - \tilde{A}_\tau(\text{pH}_i, \vec{\theta})}{\sigma_i} \right)^2 \tag{13}$$

where $A_\tau(\text{pH}_i)$ and $\tilde{A}_\tau\left(\text{pH}_i, \vec{\theta}\right)$ are the experimental and theoretical CadA activities after time τ at a given pH level pH_i ($i=1,..N_{\text{pH}}$), respectively. σ_i denotes the standard error of the measurement, and N_{pH} is the total number of data points. The other contributions, χ^2_{lys} and χ^2_{cad}, are defined equivalently to Eq. (13). However, they differ in the time at which the CadA activity was determined experimentally—that is, $\tau_{\text{pH}}=1.5$ h, $\tau_{\text{lys}}=1.5$ h, and $\tau_{\text{cad}}=3$ h.

To account for the presence of local optima and to quantify the uncertainty in the estimated parameters, we performed 1000 independent fits with randomly chosen initial parameter sets (within their physiological ranges). In Fig. 5, the final χ^2 values are plotted against the final parameters. We followed Wall *et al.*[49] to compute the asymmetric errors σ_+ and σ_- with respect to the optimal parameter values $\vec{\theta}^{\text{opt}}$ listed in Table 2. The squared errors for parameter θ_k were calculated using the following equations:

$$\sigma^2_{k,+} = \frac{\sum_{i:\theta_{k,i}>\theta_k^{\text{opt}}} \left(\theta_{k,i} - \theta_k^{\text{opt}}\right)^2 e^{-\chi_i^2/2}}{\sum_{i:\theta_{k,i}>\theta_k^{\text{opt}}} e^{-\chi_i^2/2}} \quad \text{and}$$
$$\sigma^2_{k,-} = \frac{\sum_{i:\theta_{k,i}<\theta_k^{\text{opt}}} \left(\theta_{k,i} - \theta_k^{\text{opt}}\right)^2 e^{-\chi_i^2/2}}{\sum_{i:\theta_{k,i}<\theta_k^{\text{opt}}} e^{-\chi_i^2/2}} \tag{14}$$

where $\sigma_{k,i}$ is the value of parameter θ_k in the ith fit, θ_k^{opt} is the value of θ_k in the fit with the lowest value of χ^2, and χ_i^2 is the value of χ^2 for the ith fit. In using the likelihood function $e^{-\chi^2/2}$, we assume that the errors in the measurements are independent and normally distributed with widths equal to the standard error of the mean.

Acknowledgements

It is a pleasure to thank T. Hwa and M. Saier for helpful discussions. This work was supported by the Excellence Cluster "Nanosystems Initiative Munich"

and by Deutsche Forschungsgemeinschaft through grants JU270/5-3 and Exc114/1. I. H. was supported by a fellowship from Elite Netzwerk Bayern.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.08.037

References

1. Gale, E. F. & Epps, H. M. R. (1942). The effect of the pH of the medium during growth on the enzymic activities of bacteria (*Escherichia coli* and *Micrococcus lysodeikticus*) and the biological significance of the changes produced. *Biochem. J.* **36**, 600–618.
2. Booth, I. R. (1985). Regulation of cytoplasmic pH in bacteria. *Microbiol. Rev.* **49**, 359–378.
3. Park, Y.-K., Bearson, B., Bang, S. H., Bang, I. S. & Foster, J. W. (1996). Internal pH crisis, lysine decarboxylase and the acid tolerance response of *Salmonella typhimurium*. *Mol. Microbiol.* **20**, 605–611.
4. Richard, H. & Foster, J. W. (2004). *Escherichia coli* glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential. *J. Bacteriol.* **186**, 6032–6041.
5. Castanie-Cornet, M.-P., Penfound, T. A., Smith, D., Elliott, J. F. & Foster, J. W. (1999). Control of acid resistance in *Escherichia coli*. *J. Bacteriol.* **181**, 3525–3535.
6. Booth, I. R., Cash, P. & O'Byrne, C. (2002). Sensing and adapting to acid stress. *Antonie van Leeuwenhoek*, **81**, 33–42.
7. Merrell, D. S. & Camilli, A. (2002). Acid tolerance of gastrointestinal pathogens. *Curr. Opin. Microbiol.* **5**, 51–55.
8. Foster, J. W. (2004). *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nat. Rev. Microbiol.* **2**, 898–907.
9. Popkin, P. S. & Maas, W. K. (1980). *Escherichia coli* regulatory mutation affecting lysine transport and lysine decarboxylase. *J. Bacteriol.* **141**, 485–492.
10. Meng, S. Y. & Bennett, G. N. (1992). Nucleotide sequence of the *Escherichia coli cad* operon: a system for neutralization of low extracellular pH. *J. Bacteriol.* **174**, 2659–2669.
11. Meng, S. Y. & Bennett, G. N. (1992). Regulation of the *Escherichia coli cad* operon: location of a site required for acid induction. *J. Bacteriol.* **174**, 2670–2678.
12. Neely, M. N., Dell, C. L. & Olson, E. R. (1994). Roles of LysP and CadC in mediating the lysine requirement for acid induction of the *Escherichia coli cad* operon. *J. Bacteriol.* **176**, 3278–3285.
13. Neely, M. N. & Olson, E. R. (1996). Kinetics of expression of the *Escherichia coli cad* operon as a function of pH and lysine. *J. Bacteriol.* **178**, 5522–5528.
14. Torres, A. G. (2009). The *cad* locus of *Enterobacteriaceae*: more than just lysine decarboxylation. *Anaerobe*, **15**, 1–6.
15. Moreau, P. L. (2007). The lysine decarboxylase CadA protects *Escherichia coli* starved of phosphate against fermentation acids. *J. Bacteriol.* **189**, 2249–2261.
16. Sabo, D. L., Boeker, E. A., Byers, B., Waron, H. & Fischer, E. H. (1974). Purification and physical properties of inducible *Escherichia coli* lysine decarboxylase. *Biochemistry*, **13**, 662–670.
17. Soksawatmaekhin, W., Kuraishi, A., Sakata, K., Kashiwagi, K. & Igarashi, K. (2004). Excretion and uptake of cadaverine by CadB and its physiological functions in *Escherichia coli*. *Mol. Microbiol.* **51**, 1401–1412.
18. Watson, N., Dunyak, D. S., Rosey, E. L., Slonczewski, J. L. & Olson, E. R. (1992). Identification of elements involved in transcriptional regulation of the *Escherichia coli cad* operon by external pH. *J. Bacteriol.* **174**, 530–540.
19. Dell, C. L., Neely, M. N. & Olson, E. R. (1994). Altered pH and lysine signalling mutants of *cadC*, a gene encoding a membrane-bound transcriptional activator of the *Escherichia coli cadBA* operon. *Mol. Microbiol.* **14**, 7–16.
20. Küper, C. & Jung, K. (2005). CadC-mediated activation of the *cadBA* promoter in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* **10**, 26–39.
21. Pfau, J. D. & Taylor, R. K. (1998). Mutations in *toxR* and *toxS* that separate transcriptional activation from DNA binding at the cholera toxin gene promoters. *J. Bacteriol.* **180**, 4724–4733.
22. Laub, M. T. & Goulian, M. (2007). Specificity in two-component signal transduction pathways. *Annu. Rev. Genet.* **41**, 121–145.
23. Tetsch, L., Koller, C., Haneburger, I. & Jung, K. (2008). The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Mol. Microbiol.* **67**, 570–583.
24. Lee, Y. H., Kim, J. H., Bang, I. S. & Park, Y. K. (2008). The membrane-bound transcriptional regulator CadC is activated by proteolytic cleavage in response to acid stress. *J. Bacteriol.* **190**, 5120–5126.
25. Jung, K. & Altendorf, K. (2002). Towards an understanding of the molecular mechanisms of stimulus perception and signal transduction by the KdpD/KdpE system of *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* **4**, 223–228.
26. Klipp, E., Nordlander, B., Kruger, R., Gennemark, P. & Hohmann, S. (2005). Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.* **23**, 975–982.
27. Wahde, M. & Hertz, J. (2000). Coarse-grained reverse engineering of genetic regulatory networks. *BioSystems*, **55**, 129–136.
28. Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. & Palsson, B.Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* **7**, 129–143.
29. Sourjik, V. & Berg, H. C. (2002). Receptor sensitivity in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA*, **99**, 123–127.
30. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J. & Phillips, R. (2005). Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* **15**, 116–124.
31. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J. *et al.* (2005). Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* **15**, 125–135.
32. Lutz, R. & Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1–I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210.
33. Kuhlman, T., Zhang, Z., Saier, M. H., Jr & Hwa, T. (2007). Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **104**, 6043–6048.

34. Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 9697–9702.
35. Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R. & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, 1871–1878.
36. Selinger, D. W. & Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* **13**, 216–223.
37. Rosen, B. P. (1971). Basic amino acid transport in *Escherichia coli*. *J. Biol. Chem.* **246**, 3653–3662.
38. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M. *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
39. Tabor, H., Hafner, E. W. & Tabor, C. W. (1980). Construction of an *Escherichia coli* strain unable to synthesize putrescine, spermidine, or cadaverine: characterization of two genes controlling lysine decarboxylase. *J. Bacteriol.* **144**, 952–956.
40. Miller, J. H. (1992). *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli* and Related Bacteria. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
41. Epstein, W. & Kim, B. S. (1971). Potassium transport loci in *Escherichia coli* K-12. *J. Bacteriol.* **108**, 639–644.
42. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
43. Lemonnier, M. & Lane, D. (1998). Expression of the second lysine decarboxylase gene of *Escherichia coli*. *Microbiology*, **144**, 751–760.
44. Phan, A. P. H., Ngo, T. T. & Lenhoff, H. M. (1982). Spectrophotometric assay for lysine decarboxylase. *Anal. Biochem.* **20**, 193–197.
45. Aiba, H., Adhya, S. & de Crombrugghe, B. (1981). Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J. Bacteriol.* **256**, 11905–11910.
46. Jung, K., Krabusch, M. & Altendorf, K. (2001). Cs^+ induces the *kdp* operon by lowering the intracellular K^+ concentration. *J. Bacteriol.* **183**, 3800–3803.
47. Driessen, A. J., Molenaar, D. & Konings, W. N. (1989). Kinetic mechanism and specificity of the arginine–ornithine antiporter of *Lactococcus lactis*. *J. Biol. Chem.* **264**, 10361–10370.
48. Koppelman, R., Mandeles, S. & Hanke, M. E. (1958). Use of enzymes and radiocarbon in estimation of the equilibrium constants for the decarboxylation of lysine and glutamate. *J. Biol. Chem.* **230**, 73–80.
49. Wall, M. E., Markowitz, D. A., Rosner, J. L., & Martin, R. G. (2009). Model of transcriptional activation by MarA in *Escherichia coli*. arXiv:0902.0959v1.

Supplementary Material to: Induction kinetics of a conditional pH stress response system in *Escherichia coli*

Georg Fritz [a], Christiane Koller [b], Korinna Burdack [b], Larissa Tetsch [b], Ina Haneburger [b], Kirsten Jung [b], and Ulrich Gerland [a,*]

[a] *Arnold-Sommerfeld Center for Theoretical Physics, LMU München, Theresienstr. 37, 80333 München, Germany*

[b] *Munich Center of integrated Protein Science CiPSM at the Department of Biology I, Microbiology, LMU München, Großhaderner Straße 2-4, 82152 Martinsried, Germany*

[*] *Corresponding author, Email address:* gerland@lmu.de

Correlation analysis between parameters

From Fig. 5 in the main paper we found that for small values of χ^2 some parameters still display a wide variation over their physiological range, as e.g. the transcription and translation rate ν_m and ν_p, respectively. Therefore we reasoned, whether these 'sloppy' parameters do not influence the model behavior at all, or whether not single parameters, but rather combinations of those parameters are constrained by our experiments. To discriminate between these alternative explanations, we selected the best fits with $\chi^2 < 3.22$ from Fig. 5 in the main paper and analyzed the correlations between the fitted parameters, see Fig. S1. This data clearly displays strong pairwise correlations between the sloppy parameters, demonstrating that in fact only their combinations are well constrained. For instance, the transcription and translation rates ν_m and ν_p, which display high variation individually, are strongly anticorrelated, such that their product is kept at a constant level. This is reasonable, since the absolute level of the CadA activity - characterized by $\nu_m \times \nu_p$ - is well determined in our experiments, whereas the mRNA level - characterized by ν_m alone - is only measured in relative units. Likewise, the parameters for lysine turnover, v_{max} and K_m, display a linearly correlated variation, indicating that only their ratio is well constrained by our data. Also among the parameters that control the promoter activity function (K_m/C_0, the promoter fold-change f, K_c and

K_l; cf. Eqs. (2) and (3) in the main paper) pairwise correlations, albeit less pronounced, were detected.

Comparison of estimated parameters with literature values

In the following we discuss the estimated parameters of the Cad module listed in Table 2 of the main paper in the light of previously published literature values.

$\mathsf{K_l}$: activation threshold of CadC by lysine. CadC senses lysine indirectly via interaction with LysP [1–4]. There is recent evidence that CadC is inhibited at low lysine concentrations via a transmembrane domain interaction with the lysine permease LysP, whereas the interaction is released at high lysine levels [5]. In the present work we determined the effective *in vivo* lysine activation threshold to be $K_l = 3.6\,\mathrm{mM}$. This result is somewhat surprising, since the *in vitro* Michaelis constant K_M for lysine transport by LysP is much lower at $\sim 10\,\mu\mathrm{M}$ [6]. However, Neely *et al.* also found that the Cad system is sensitive with respect to lysine in the millimolar range [3], suggesting that our result is plausible under *in vivo* conditions.

$\mathsf{n_l}$: Hill exponent of indirect CadC-lysine interaction. The estimated Hill exponent of $n_l = 1.1$ for the indirect activation of CadC by lysine via LysP suggests that the underlying binding reactions do not involve cooperative mechanisms of CadC activation. We are not aware of a literature value for the Hill exponent for this interaction to which we could compare our estimate.

$\mathsf{K_c}$: inactivation threshold of CadC by cadaverine. For the inactivation of CadC by external cadaverine we found a threshold of $K_c = 235\,\mu\mathrm{M}$. This value is surprisingly close to the *in vitro* binding constant of $96\,\mu\mathrm{M}$ for the interaction of cadaverine with the periplasmic domain of CadC [5].

$\mathsf{n_c}$: Hill exponent of direct CadC-cadaverine interaction. It is noteworthy that the best-fit value for the effective Hill coefficient n_c for the regulation of CadC by cadaverine is 2.8, and is relatively well constrained by the data. This suggests that a molecular mechanism for cooperativity is at work, possibly a multimerization of CadC proteins in the membrane. To the best of our knowledge, this question has so far not been studied *in vivo*.

$\mathsf{pH_0}$ and $\mathsf{\Delta pH}$: activation threshold of CadC by low pH. Our analysis yields for the activation threshold of CadC $\mathrm{pH}_0 = 6.2$ and for the width of the activation range $\Delta\mathrm{pH} = 0.5$. These values agree reasonably well with early experiments of Auger *et al.* in Falkow medium, who found activity of a *cadA-lacZ* fusion at pH 6 and below [7]. Also from a physiological point of view it seems reasonable to

activate the Cad system in this pH regime, since the enzyme activity of CadA is peaked around an optimal pH of 5.7 [8].

$\mathsf{C_0/K_c}$: relative CadC abundance. The level of CadC was found to be just sufficient to activate the promoter P_{cad}, such that multiple copies of a $P_{cad} - lacZ$ fusion were not fully induced [2]. This result suggests, that *in vivo* the number of CadC molecules is on the order of the dissociation constant for CadC-promoter binding, i.e., $\mathsf{C_0} \approx \mathsf{K_C}$. In our analysis we found indeed $C_0/K_c = 1.1$, which is in excellent agreement with the observations of Watson *et al.* [2].

ν_m and ν_p: transcription and effective translation rate of *cadBA*. As inferred from Fig. S1, only the product of transcription and effective translation rate is well constrained by our data. From our analysis we found $\nu_m \times \nu_p = 18\,\mathrm{U/mg/min^2}$. We are not aware of any reference values for these parameters in the literature.

f: fold-change between basal and maximal transcription rate of P_{cad}. For the fold-change we found $f \approx 700$, which indicates that the promoter is highly inducible. Again, we are not aware of any reference value in the literature.

τ_m: half-life of *cadBA* mRNA. From our analysis we found an mRNA half-life of 13.8 min. In contrast, a global analysis of RNA half-lifes in *Escherichia coli* [9] found in an extremely short half-life of less than 2 min for the *cadBA* mRNA in LB medium, suggesting that an active degradation mechanism is involved. Our Northern blot data for the *lysP211* mutant, shown in Fig. 4 (f) of the main paper, does indeed suggest a rapid decay of the mRNA at high levels directly after the peak, followed by a slower decay at lower levels. Our quantitative model only allows for a single degradation rate, which leads to the intermediate half-life of 14 min as a best-fit value. The changing degradation rate could be rationalized under the assumption that the *cadBA* mRNA has a relatively weak binding affinity to the degrading enzyme, such that active degradation only contributes significantly at high mRNA levels, whereas at low mRNA levels a slower passive decay is at work. Fig. S2 of the Supplementary Material illustrates that this mechanism could indeed account for the observed shape of the mRNA curve in Fig. 4(f). However, this explanation would raise the question why the rapid active decay is not observed in the data of the wild-type in Fig. 4(a). Possibly, the kinetics of LysP unbinding from CadC is slow (contrary to our model assumption of a rapid binding equilibrium), such that it interferes with the reception of the cadaverine signal and thereby broadens the onset of transcriptional down-regulation. Further experiments would be required to test these hypotheses.

τ_p: protein half-life. For the protein half-life in our effective model we found $\tau_p = 29\,\mathrm{h}$, indicating high protein stability. This finding is in agreement with a biochemical study which found CadA to be stable [8].

v_{max} and K_m: effective turnover rate and Michealis constant for lysine turnover by CadA and CadB. As inferred from Fig. S1, only the ratio of v_{max} and K_m is well constrained by our data: $v_{max}/K_m = 5 \times 10^{-5}\,(\mathrm{min} \times \mathrm{U/mg})^{-1}$. To the best of our knowledge, there is no report of these parameters in the literature.

References

[1] P. S. Popkin and W. K. Maas. *Escherichia coli* regulatory mutation affecting lysine transport and lysine decarboxylase. *J. Bacteriol.*, 141:485–492, 1980.

[2] N. Watson, D. S. Dunyak, E. L. Rosey, J. L. Slonczewski, and E. R. Olson. Identification of elements involved in transcriptional regulation of the *Escherichia coli cad* operon by external pH. *J. Bacteriol.*, 174:530–540, 1992.

[3] M. N. Neely, C. L. Dell, and E. R. Olson. Roles of LysP and CadC in mediating the lysine requirement for acid induction of the *Escherichia coli* cad operon. *J. Bacteriol.*, 176:3278–3285, 1994.

[4] C. L. Dell, M. N. Neely, and E. R. Olson. Altered pH and lysine signalling mutants of *cadC*, a gene encoding a membrane-bound transcriptional activator of the *Escherichia coli cadBA* operon. *Mol. Microbiol.*, 14:7–16, 1994.

[5] L. Tetsch, C. Koller, I. Haneburger, and K. Jung. The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Mol. Microbiol.*, 67:570–583, 2008.

[6] B. P. Rosen. Basic amino acid transport in *Escherichia coli*. *J. Biol. Chem.*, 246:3653–3662, 1971.

[7] E. A. Auger, K. E. Redding, T. Plumb, L. C. Childs, S.-Y. Meng, and G. N. Bennett. Construction of *lac* fusions to the inducible arginine-and lysine decarboxylase genes of *Escherichia coli* K-12. *Mol. Microbiol.*, 3:609–620, 1989.

[8] D. L. Sabo, E.A. Boeker, B. Byers, H. Waron, and E. H. Fischer. Purification and physical properties of inducible *Escherichia coli* lysine decarboxylase. *Biochemistry*, 13:662–670, 1974.

[9] D.W. Selinger and C. Rosenow. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.*, 13:216–223, 2003.

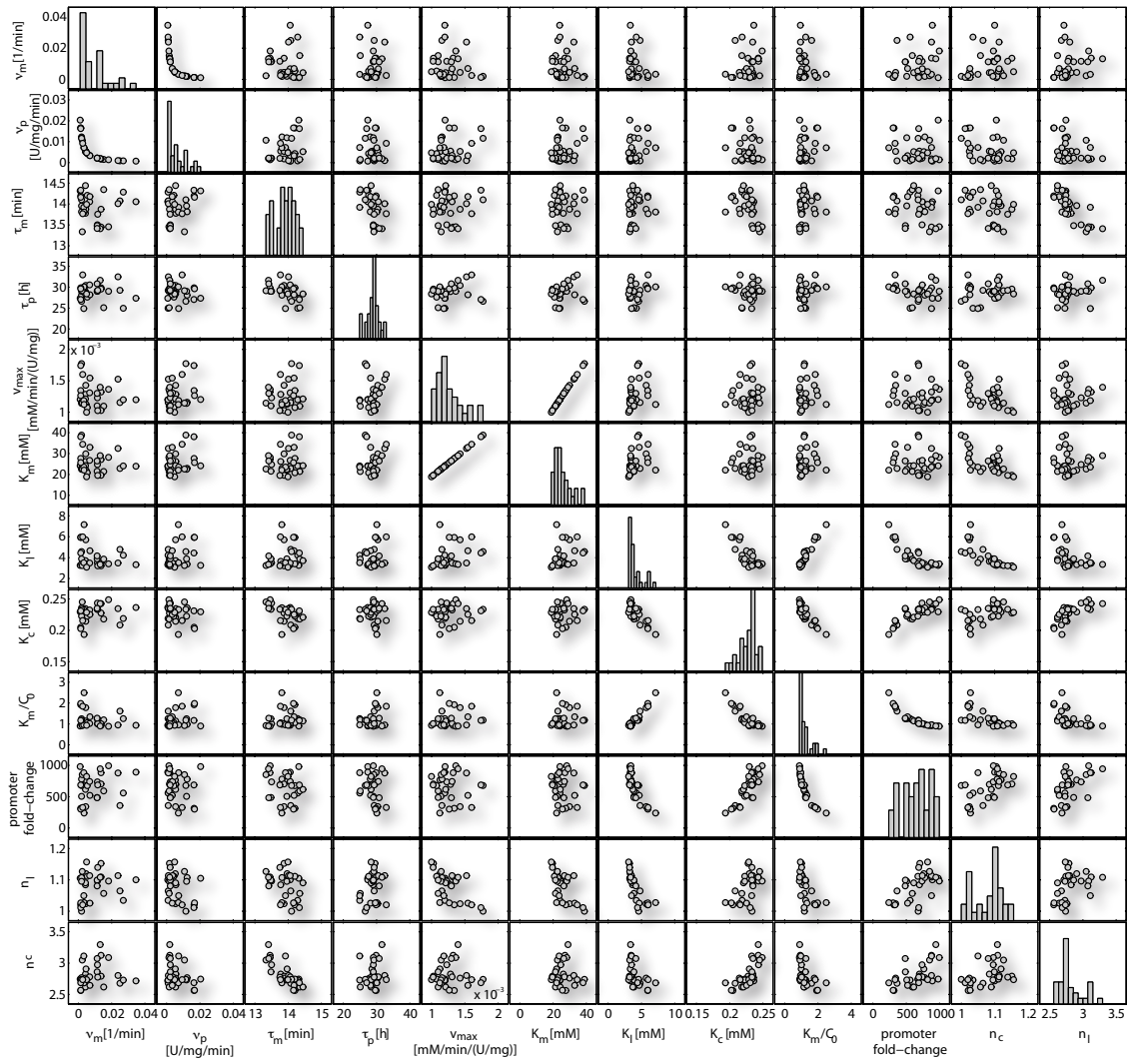


Fig. S1. Pairwise correlations between the fit parameters. For this correlation analysis, the fits of highest quality ($\chi^2 < 3.22$) were selected from Fig. 5 (main paper) and the correlations of the resulting parameters are displayed in pairwise scatterplots (*off-diagonal*). The figure also shows the histograms of the individual parameters (*diagonal*).

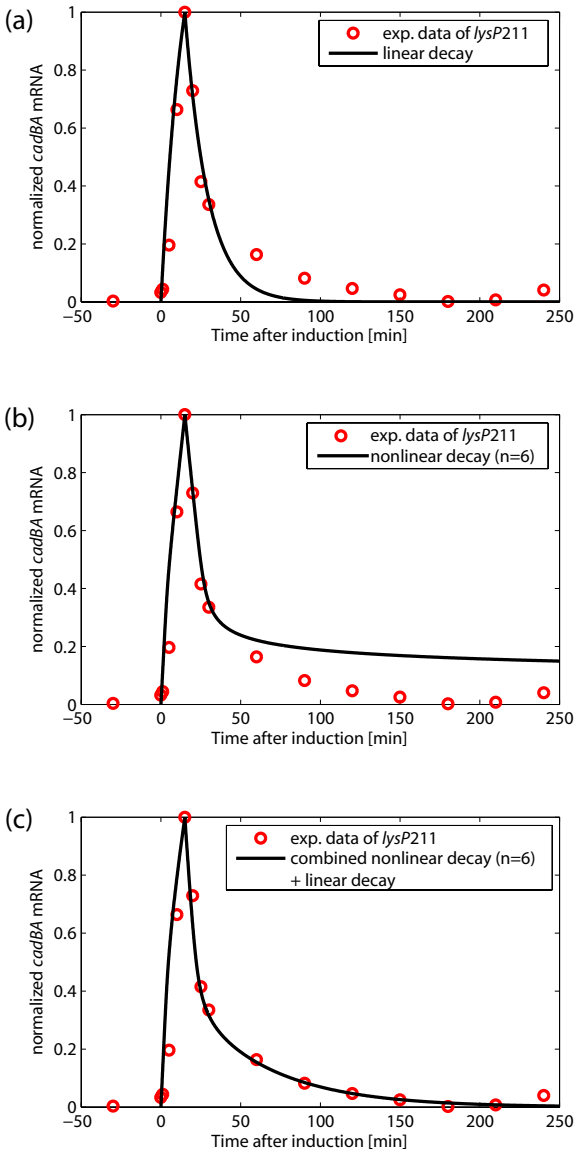


Fig. S2. Schematic influence of different mRNA degradation mechanisms on the dynamics of *cadBA* mRNA. The solid lines correspond to a model which assumes mRNA (m) production at constant rate α until $T = 15\,\text{min}$, followed by a period of pure mRNA decay: $\frac{d}{dt}m = \alpha\Theta(T-t) - g(m)$, where $\Theta(T-t)$ is the Heaviside function and $g(m)$ is the degradation term. In (a) the mRNA dynamics with simple linear decay ($g(m) = \lambda m$) shows major differences to the experimental data of the *lysP211* mutant after $t = 60\,\text{min}$. In (b) the mRNA dynamics is shown for a degradation mechanism which is active when the mRNA level exceeds a threshold value K, i.e. $g(m) = \lambda m^n/(K^n + m^n)$. The Hill exponent was chosen to be large ($n = 6$) in order to obtain a switch-like dependence of the mRNA degradation rate on the mRNA level: below $K = 0.4$ degradation turns off and hence the long-term behavior of the experimental data ($t > 60$ min) is not captured. Only a combination of nonlinear active decay and slow linear decay in (c) yields the required short *and* long-term behavior.

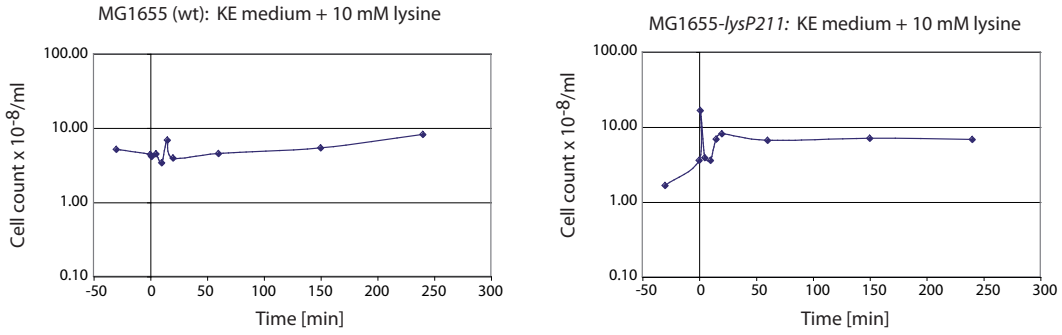


Fig. S3. Dynamics of the life cell counts of *E. coli* MG1655 (*left*) and MG1655-*lysP211* (*right*) upon exposure to pH 5.8 and 10 mM lysine in minimal KE medium.

The feedback-inhibitor cadaverine suppresses pH response by binding to the pH susceptible site of CadC in *Escherichia coli**

Ina Haneburger[‡,1,2], Georg Fritz[‡,3], Nicole Jurkschat[1,2], Larissa Tetsch[1,2], Andreas Eichinger[1,4], Arne Skerra[1,4], Ulrich Gerland[3] and Kirsten Jung[1,2]

[1]from the Center of Integrated Protein Science Munich at the [2]Department of Microbiology, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany, the [3]Arnold Sommerfeld Center for Theoretical Physics and the Center for NanoScience at the Ludwig-Maximilians-Universität München, 80333 München, Germany and the [4]Lehrstuhl für Biologische Chemie, Technische Universität München, 85350 Freising-Weihenstephan, Germany.

*Running title: *Cadaverine forces CadC into an inactive conformation*

To whom correspondence should be addressed: Kirsten Jung, Dept. Biologie I, Mikrobiologie, Biozentrum, Ludwig-Maximilians-Universität München, Grosshadernerstr. 2 - 4, 82152 Martinsried, Germany, Tel.: +49-89-2180-74500; Fax: +49-89-2180-74520; E-mail: jung@lmu.de

Keywords: acid stress; mathematical model; one-component system; polyamine; ToxR

Background: CadC is a membrane-integrated sensor and transcriptional activator that integrates various external signals.
Results: CadC is inactivated by binding of the feedback inhibitor cadaverine to the activation site.
Conclusion: It is suggested that cadaverine binding interferes with the active CadC dimer conformation.
Significance: This study provides molecular details on feedback regulation mediated by a periplasmic sensor domain.

SUMMARY

At acidic pH and in the presence of lysine the pH-sensor CadC activates transcription of the *cadBA* operon encoding the lysine-cadaverine antiporter CadB and the lysine decarboxylase CadA. In effect, these proteins convert external lysine into external cadaverine under the consumption of a cytoplasmic proton, thus contributing to acid stress adaptation in *Escherichia coli*. Earlier, it was shown that *cadBA* expression is feedback inhibited by cadaverine, and a cadaverine binding site was predicted within the central cavity of the periplasmic domain of CadC. Our present study demonstrates that the predicted binding site only partially accounts for the cadaverine sensitivity of CadC *in vivo*. Moreover, a second, pivotal binding site was identified, which is localized at the pH-responsive patch of amino acids at the surface of the periplasmic domain of CadC, directly at the homo-dimerization interface. These data were incorporated into a previously established mathematical model for the Cad module, and the temporal response upon acid shock was simulated for two cadaverine-insensitive CadC variants. Based on these data a model of sequential binding of cadaverine to CadC is suggested. Binding of cadaverine within the central cavity renders the second site accessible for cadaverine, and binding of cadaverine finally leads to the inactivation of CadC. Altogether, the data reveal binding of a feedback inhibitor directly to the activation site of a receptor, and represent a striking example for the deactivation of a pH-sensor.

Rapid adaptation to changing environmental conditions is pivotal for the reproductive success of microbes. For instance, enterobacteria, such as *E. coli*, thrive in- and outside the human host, whereby growth conditions change suddenly and deviate drastically from their optimal ranges (1). Adaptation to these ever-changing conditions requires accurate monitoring of critical parameters and a precise and specific extra/intracellular information flow in order to respond with appropriate alterations in gene expression and protein activity (2). Therefore, it is still challenging to understand how stress response

systems integrate and process multiple input signals and generate conditional responses to environmental stimuli.

One such conditional stress response system is the Cad module of *E. coli*, in which the one-component receptor system CadC detects external acidification and activates transcription of the *cadBA* operon (3). *cadA* encodes the inducible lysine decarboxylase and converts lysine under consumption of a cytoplasmic proton to cadaverine and carbon dioxide (4). Cadaverine is subsequently excreted via the lysine/cadaverine antiporter CadB. During this cycle, the proton consumption in the cytoplasm and the export of alkaline cadaverine are believed to be the main mechanisms to tackle acid stress (5). Since induction of *cadBA* under stress conditions (acidic pH, low oxygen tension) results in a drastic increase of the corresponding proteins [CadA increases to approx. 2% of cellular protein, see ref. (6)], transcription has to be tightly controlled. In fact, it is known that the Cad module is only activated if in addition to the primary stimulus (low pH), the substrate for the decarboxylase is also present (high lysine level) (7) and the product of the reaction is absent (low cadaverine level) (8,9). Hence, CadC unifies the remarkable abilities of integrating and processing three external cues, transmitting this information to the cytoplasm and, accordingly, activating transcription of *cadBA*. Activation in the presence of lysine is most likely accomplished by removal of an inhibitory interaction of CadC with the lysine permease LysP mediated by the transmembrane domain (10). Acidification of the external milieu is most probably sensed by protonation of a patch of acidic amino acids at the dimer interface within the periplasmic domain of CadC (12). This neutralization enables conformational changes that lead to activation of *cadBA* transcription (11). While these studies have shed light on the mechanisms of lysine sensing and low pH detection, the molecular details of feedback inhibition of CadC by cadaverine are still unclear.

Crystal structure analysis of the periplasmic sensor domain of CadC ($CadC_{pd}$) revealed a central cavity between the two periplasmic subdomains of each CadC monomer, which is adequate for accommodation of the positively charged cadaverine - both with respect to sterical fit and electrostatic complementarity (12).

In the present study we functionally analyze the molecular details of CadC inactivation by the feedback inhibitor cadaverine. We identify a second, pivotal cadaverine binding site that overlaps with the patch of pH-responsive amino acids. Mathematical simulations support a sequential inactivation model for CadC. It is proposed that binding of cadaverine to the first site induces conformational changes which enable accessibility to the second binding site, and rapid inactivation of CadC.

EXPERIMENTAL PROCEDURES

Bacterial strains and growth conditions - *E. coli* strains JM109 or DH5α (13,14) were used as cloning hosts for the plasmids listed in Table S1. *E. coli* EP314 (9), which carries a *cadA'–lacZ* fusion gene and a deletion in *cadC*, was complemented with plasmids (pET16b-based) encoding *cadC* and its variants and was used to determine *cadBA* expression. For strain propagation, plasmid preparation and protein overproduction, strains were grown on Luria Bertani (LB) medium (15). To monitor signal transduction *in vivo*, *E. coli* EP314 transformed with the indicated plasmids was grown in minimal medium as described below and previously published (10,16). Antibiotics were added for selection at concentrations of 100 µg ml^{-1} (ampicillin), 50 µg ml^{-1} (kanamycin) and 34 µg ml^{-1} (chloramphenicol). For monitoring *cadBA* expression as well as CadA activity and cadaverine accumulation MG1655Δ*cadC* (MG1655 *cadC::kan*; MG1655 (17) x P1[W3110-Δ*cadC* (18)]) was transformed with plasmids pET16b-cadC2 (10), pET16b-cadC_Y453I (11) and pET16b-cadC_T475A, respectively. These three strains were then cultivated as described earlier (8).

Measurement of extracellular cadaverine, intracellular CadA activity and cadBA *expression* - Enzymatic activity, cadaverine content and *cadBA* transcription were determined as described earlier (8).

Construction of cadC *variants* – Mutagenesis was performed by either one- or two-step PCR using the appropriate oligonucleotides (Table S1). To facilitate construction, *cadC* genes harboring singular restriction sites were employed (10,11).

In vivo *analysis of cadaverine dependent inhibition* - Cultivation of cells in minimal

medium with 10 mM lysine and measurements of signal transduction activity of different CadC variants by ß-galactosidase assays were performed as described before (10). In addition, cells were cultivated with addition of cadaverine/HCl at the concentrations indicated to test for the effect on CadC-dependent *cadBA* expression.

Determination of in vitro *affinity for cadaverine* – The periplasmic domains of CadC_T229A_E447Q and wild type CadC were produced and purified as described in (10). Purification was complemented by an additional size exclusion chromatography step (Superdex 200 10/300 GL, Amersham Pharmacia), to remove the thioredoxin cleavage product of the initial fusion protein (10). The purified protein was used for scintillation proximity assays (with ^{14}C-cadaverine from Hartmann analytic) according to (19) or tryptophan fluorescence measurements as described in (10) with a slightly changed buffer (Trp-buffer: 50 mM sodium phosphate buffer (pH 7.0) and 150 mM NaCl).

Computational Model - To analyze the *in vivo* dynamics of the Cad module at a quantitative level, we previously established a mathematical model for the dynamics of the three system variables *cadBA* mRNA, CadA activity, and external cadaverine monitored in our experiments (8). Briefly, the model was guided by the known biochemistry of the Cad module, describing the integration of the input signals by CadC, a quasi-static CadC activity function and accounting for the dynamics of transcription, translation and enzyme kinetics of the Cad module via ordinary differential equations. Here we have refined this model by including the perception that CadC is active as a dimer (12). The fraction of active CadC dimer per cell ($CadC_2$) at time t ($C_2(t)$), with respect to the maximal number of $CadC_2$ dimer per cell (C_0) is defined as a function of its input signals - pH*(t)*, external lysine, *l(t)*, and external cadaverine, *c(t)* - and now reads

$$C_2(t)/C_0 = f(pH(t)) \times g(l(t)) \times h(c(t)) \quad (1),$$

wherein the functions *f*, *g*, and *h* in Eq. (1) take on values between 0 and 1 and are assumed to be of the Hill form typical for cooperative binding reactions. For instance, the cadaverine dependence takes the form

$$h(c(t)) = \frac{1}{1+\left(c(t)/K_c\right)^{n_c}} \quad (2),$$

wherein K_c is the apparent cadaverine affinity of $CadC_2$ and n_c the corresponding Hill exponent [see (8) for all details of the other regulation functions]. Transcriptional regulation of P_{Cad} by $CadC_2$ is described by a quantitative thermodynamic model analogous to (8) and the resulting rate equation for the time evolution of *cadBA* mRNA, *m(t)*, takes the form

$$\frac{d}{dt}m(t) = \nu_m \left(\frac{1+\omega\left(C_2/K_{C_2}\right)}{1+C_2/K_{C_2}} \right)^2 - \lambda_m m(t) \quad (3).$$

Here, ν_m is the basal transcription rate of the *cadBA* promoter (P_{Cad}), ω the fold change between basal and maximal transcription rate, K_{C2} the affinity of a single $CadC_2$ binding site on the *cadBA* promoter (P_{Cad} harbors two binding sites (21) – presumably for the dimer form of CadC – leading to the square in the first term of the right hand side) and λ_m the *cadBA* mRNA degradation rate. The kinetics of enzyme expression, conversion of lysine to cadaverine and subsequent antiport by CadB were modeled as before (8). Again, parameters were estimated by least-squares minimization. Briefly, all previously published data (8) were fitted with the refined model described above (the full list of fit parameters is given in Table S2). To account for the slightly faster Cad response due to the somewhat higher copy number of the plasmid-encoded CadC used here, compared to strain MG1655 (8), the transcription rate of P_{Cad} was varied as the only fit parameter (Table S2). Subsequently, the dynamical responses of the two mutant strains harboring CadC_T229A_E447Q and CadC_Y453I were fitted by varying only the cadaverine affinity K_c, the Hill exponent n_c and the maximal number of $CadC_2$ dimers C_0 (see Fig. S2 and S3 for the fit results).

RESULTS

Cadaverine binding within the internal cavity – The product of lysine decarboxylation, cadaverine, has been shown to inhibit expression of the *cadBA* operon (9). Direct binding of cadaverine to the periplasmic domain of CadC was determined *in vitro* with an apparent K_d of 96 µM (10). The crystal structure of $CadC_{pd}$ was recently solved (12). As crystallization in the presence of cadaverine did not result in a $CadC_{pd}$ complex structure, docking studies with cadaverine were performed. The ten best scoring results of the docking studies performed earlier (12) proposed a binding mode of cadaverine with an extended conformation within the central cavity of the periplasmic domain of CadC. According to these findings one amine group of cadaverine was coordinated by Glu447, Gln421 and Tyr374 and the second amine group by Asp225 and Thr229. To verify the results of these docking studies, we constructed CadC variants with single, double and triple amino acid substitutions at these positions. In addition, we analyzed the residues Trp450 and His344, found to coordinate the hexachlororhenate ion in the CadC complex that had been prepared for MAD phasing (12).

Cadaverine-dependent down-regulation of transcription was determined in reporter assays. *E. coli* strain EP314 lacks the native *cadC* gene and possesses a *cadA'-lacZ* fusion gene. Therefore, this strain is unable to produce cadaverine via CadA (as a large part of *cadA* is replaced by *lacZ*) and, hence, cadaverine-dependent inhibition only relies on the addition of external cadaverine while CadC-dependent *cadBA* promoter activity can be monitored via ß-galactosidase measurements. Complementation of this strain with plasmid-encoded CadC or CadC variants (pET16b-based, see Experimental Procedures) allows the survey of CadC-dependent *cadBA* expression depending on externally applied stimuli. A spectrum of cadaverine concentrations (0 mM; 0.65 mM; 1.3 mM; 2.6 mM; 5.2 mM; 10.4 mM) was tested to determine the inhibitory effects on *cadBA* expression. Increasing cadaverine concentrations inhibited wild type CadC in a dose-dependent manner (Fig. 1A). Approximately 1 mM cadaverine was required to inhibit CadC wild type-dependent *cadBA* expression to about 50% (IC_{50}) (Table 1). In the presence of 10.4 mM cadaverine expression was inhibited by nearly 100 % (residual activity in the presence of 10.4 mM cadaverine: 1%) (Fig. 1A). Most of the variants (CadC_D225A, CadC_D225N, CadC_Y374A, CadC_Y374F, CadC_W450A, CadC_W450F, CadC_Q421A, CadC_D225N_T229A_E447Q, CadC_H344L, CadC_H344L_E447Q, CadC_T229A_H344L) with substitutions in the central cavity were inhibited by cadaverine like wild type CadC (IC_{50} < 2 mM; Fig. S1). Only the variants CadC_D225W, CadC_T229A and CadC_E447Q as well as the variants with multiple amino acid substitutions, CadC_D225N_E447Q, CadC_ T229A_E447Q, CadC_T229A_H344L _E447Q and CadC_T229A_Q421A_E447Q were inhibited to a lesser extend by cadaverine (Fig. 1A). In addition, the IC_{50} of those variants impaired in cadaverine sensing was significantly increased (Table 1). Polyamine binding often occurs at acidic or aromatic residues (22-26). Therefore, several substitutions at positions Asp225, Tyr374 and Trp450 were analyzed (D225A, D225N, D225W; Y374A, Y374F; W450A, W450F). Replacement by alanine, asparagine or phenylalanine did not affect cadaverine sensing, compromising the idea that Asp225, Tyr374 and Trp450 directly coordinate cadaverine. Nonetheless, replacement by the more bulky tryptophan at position Asp225 reduced the inhibitory effect of cadaverine (Table 1, IC_{50}: 3.4 mM cadaverine). Altogether, these experimental data indicated that the central cavity is involved in cadaverine sensing, as certain side chain substitutions within the cavity clearly reduced the inhibitory effect of cadaverine on CadC-dependent *cadBA* expression, although the precise nature of cadaverine coordination within the cavity remains elusive.

Identification of additional cadaverine coordinating residues – As no CadC variant with side chain substitutions within the central cavity was completely impaired in cadaverine sensing, we searched for additional cadaverine coordinating residues. In a random mutagenesis attempt Dell *et al*. (27) identified amino acid substitutions that resulted in pH-independent CadC variants that were not affected by the addition of 2 mM cadaverine. Remarkably, variants CadC_T475A and CadC_L479S, both impaired in cadaverine-dependent inhibition, carry substitutions that co-localize with the patch of acidic amino acids in the periplasmic domain

which was previously identified to be crucial for detection of acidification (11). Consequently, we examined whether cadaverine sensing and pH detection are linked to each other.

To test this hypothesis, we investigated the effect of increasing cadaverine concentration on *cadBA* expression regulated by CadC variants harboring substitutions close to the patch of acidic amino acids. For this purpose, we used CadC variants that were previously constructed to analyze pH-dependent signaling (11). CadC_H240L, CadC_D471A, CadC_D471N, CadC_E468D and CadC_F477A were strongly affected by increasing concentrations of cadaverine similar to wild type CadC (Fig. 1B, Table 1). CadC_N478A, CadC_F477I and CadC_Y453F were less affected by cadaverine (Fig. 1C, Table 1). In contrast, cadaverine exerted nearly no inhibiting effect on CadC_Y453I, CadC_Y453A, CadC_F477I and CadC_L474A (Fig. 1C, Table 1). As the activity of these variants in the presence of 10.4 mM cadaverine still exceeded 60%, the IC_{50} can be assumed to be above 10 mM (Table 1).

To understand whether the inability of these variants to sense cadaverine was evoked just by replacement of the amino acid side chain or by indirect effects such as conformational changes, Tyr453 was exchanged by phenylalanine (to preserve the aromatic character), isoleucine (insertion of an aliphatic non-aromatic side chain) or alanine (to remove nearly the complete side chain). Our data reveal that CadC is inactivated in the presence of cadaverine as long as an aromatic side chain is present at position 453 (wt and Y453F) (Fig. 1C). However, cadaverine can no longer exert its inhibitory effect if the aromatic side chain is removed either through substitution by a short (Ala) or by an aliphatic side chain (Ile).

For residue Thr475 the influence of a substitution by serine was tested. Since the variant CadC_T475S was not inactivated by cadaverine (cadaverine did not inhibit *cadBA* expression in a strain harboring this variant as well as CadC_T475A), it is likely that for Thr475 the nature of the branched amphipathic side chain is crucial. Since CadC_F477A is inhibited by cadaverine, the effect seen with CadC_F477I seems not to be caused by the lack of the aromatic side chain. It seems likely that the branched aliphatic side chain of isoleucine interferes with the 3D structure of CadC, thereby impairing signal sensing. CadC_F477I reportedly is also impaired in pH detection (11), as it induces *cadBA* expression irrespective of the external pH. Together with the results for CadC_T475A and those for CadC_L479A from Dell *et al.* (27) these data suggest a second cadaverine binding site close to the patch of acidic amino acids at the dimer interface encompassing residues Tyr453, Thr475, Leu474 and Leu479, since CadC variants with substitutions at these positions activate expression of *cadBA* irrespective of the presence of cadaverine. Closer inspection of the three-dimensional structure of CadC suggested that cadaverine binding to the second binding site might arise from a small number of residues at the dimer interface (Fig. 2, *red residues*). It appears that there are two clearly defined and separate binding sites at the dimer interface; one on each monomer. Although cadaverine binding and pH detection seem to be mediated by the same protein region, it is likely that they are independent of each other, as some variants are only affected in one of the two functions: CadC_Y453I induces *cadBA* expression in response to acid stress but is not inhibited by cadaverine; in contrast, CadC_D471A and CadC_D471N activate *cadBA* expression irrespective of the external pH (11) and are inhibited by cadaverine similarly to wild type.

Probing in vitro *affinity* – We intended to test the *in vitro* affinity of the periplasmic domains of distinct cadaverine insensitive variants [CadC_T229A_E447Q, CadC_T475A and CadC_Y453I (11)] for cadaverine. However, Trx-$CadC_{pd}$_Y453I and Trx-$CadC_{pd}$_T475A were produced in inclusion bodies and could not be used for the assay. As both variants contain substitutions at the dimer interface, this could indicate that substitutions at the dimer interface lead to alterations in the protein conformation that render the protein more insoluble. For CadC_T475A, both a tryptophan fluorescence study (10) as well as a scintillation proximity assay failed, since we measured a linear dependency between cadaverine concentrations and the corresponding signals. Thus, it can be assumed that affinity for cadaverine is greatly reduced in this variant.

Dynamics of the Cad module of CadC variants impaired in cadaverine sensing - In order to infer key biochemical parameters of the Cad

module *in vivo*, we previously established a quantitative mathematical model (8). The model parameters were calibrated with a comprehensive set of experimental data, and the parameter-free predictions for the expression behavior under various initial conditions and for a LysP mutant deficient in lysine sensing were confirmed experimentally (8). Likewise, here we reasoned that two CadC variants impaired in cadaverine sensing (CadC_T229A_E447Q and CadC_ Y453I) should produce a significantly altered response after a shift to low pH. We used our model to interpret the experimental changes in the dynamics of the Cad module.

First we refined our previously published model (8), by incorporating the result of Eichinger *et al.* (12) that CadC is active as a dimer ($CadC_2$; see also Experimental Procedures). Then we re-analyzed all data previously collected for the wild type Cad module (8) and extracted the *in vivo* parameters for the interaction between $CadC_2$ and its input signals H^+, lysine and cadaverine (the full list of fit-parameters is given in Table S2). Specifically, for the interaction of wild type CadC with cadaverine we found an apparent *in vivo* affinity of $K_c = 170$ µM and an effective Hill exponent $n_c = 3.5$. As the Hill exponent is a measure for the cooperativity involved in the interaction between ligand and protein, the obtained value pointed to (at least) four cadaverine binding sites within one $CadC_2$ dimer (28). This would be consistent with the idea that each CadC monomer has one binding site in its central cavity, and one at the patch of acidic amino acids (at the dimer interface). Similarly, the deduced affinity parameter $K_c = 170$ µM of cadaverine for $CadC_2$ was in good agreement with the measured *in vitro* affinity [see above and (10)].

Next, we experimentally characterized the influence of the proposed binding site in the central cavity of CadC on the dynamics of the Cad module. To this end, we shifted the strain harboring CadC_T229A_E447Q from pH 7.6 to a lysine-rich medium (10 mM) with pH 5.8 and recorded the response of *cadBA* mRNA, specific CadA activity and external cadaverine level as a function of time (Fig. 3A-C; *green circles*). CadA activity and cadaverine concentration was only determined for selected time points, as it is known from earlier studies that CadA activity reaches a plateau after about 60 min and cadaverine increases almost linearly (8). As a control, we applied the same protocol to a strain with identical genetic background harboring wild type CadC (Fig. 3A-C; *black circles*; see Experimental Procedures for all details). In this case [see also (8)] *cadBA* expression was quickly turned on (Fig. 3A; *black circles*) and as a consequence, both the CadA activity (Fig. 3B, *black circles*) as well as the cadaverine concentration in the medium increased (Fig. 3C; *black circles*). Once external cadaverine reached a certain threshold, transcription of *cadBA* was turned off again, thus leading to a transient expression of *cadBA*. Qualitatively, the CadC_T229A_E447Q harboring strain displayed a similar transient induction, but with a significantly widened expression peak of *cadBA* (Fig. 3A; *green circles*), a stronger increase in CadA activity (Fig. 3B; *green circles*) and a faster accumulation of cadaverine (Fig. 3C; *green circles*). These results clearly demonstrated that the substitutions in CadC_T229A_E447Q had impact on the dynamics of the Cad module *in vivo*, indicating that the targeted residues are involved in cadaverine sensing.

In order to extract more quantitative information from the dynamics of the Cad module, we leveraged our mathematical model to infer the key biochemical parameters relevant for cadaverine binding to CadC. Specifically, we used the binding constant of cadaverine for CadC, K_c, the Hill exponent, n_c, and the maximal number of active $CadC_2$ dimers, C_0, as fit parameters for the CadC_T229A_E447Q harboring strain and set all other model parameters to those of the wild type Cad module, see Experimental Procedures for details. C_0 was included as fit parameter since the dimerization of CadC might be affected by amino acid substitutions. The fit results (Fig. 3D-F; *green lines*) displayed good quantitative agreement with the experimental data (Fig. 3A-C; *green circles*), capturing both the widened expression peak of *cadBA* mRNA as well as the stronger increase in CadA activity and cadaverine level when compared to the dynamics of the wild type strain. Surprisingly, the Hill exponent was the only parameter that changed significantly between the wild type ($n_c = 3.5$) and CadC_T229A_E447Q ($n_c = 0.5$), indicating that the residues in the central cavity of $CadC_{pd}$ mainly affect the cooperativity of the interaction between cadaverine and CadC. For better illustration, the impact of these substitutions

on the $CadC_2$ activity is visualized in Fig. 4 (*green line*) as a function of cadaverine, where it is compared to wild type $CadC_2$ activity (*black line*). While the activity of latter sharply decreased with increasing cadaverine levels at K_c, the activity of $(CadC\_T229A\_E447Q)_2$ displayed a significantly reduced cadaverine concentration dependancy. Hill exponents smaller than one may indicate a negative cooperative interaction among the remaining binding sites, but also other mechnisms can be envisioned (see Discussion). Another interesting observation is that in our fit also the maximal number of transcriptionally active CadC_T229A_E447Q dimers, C_0, was increased by 30%. Indeed, the output of transcription activity (ß-galactosidase activity in Miller Units) for CadC_T229A_E447Q in our *in vivo* analyses of cadaverine-dependent inhibition was also 10% higher than for the wild type (data not shown).

Next, we analyzed the *in vivo* dynamics of a strain harboring CadC_Y453I after a shift to pH 5.8 and adding 10 mM lysine (Fig. 3A-C; *red circles*). In this strain the negative feedback via cadaverine should be nearly abolished (Fig. 1C). Nevertheless, we determined a transient *cadBA* expression with a significantly slower decrease in comparison to the strains harboring wild type CadC or CadC_T229A_E447Q (Fig. 3A; *red circles)*. However, despite this significantly widened expression peak both the specific CadA activity (Fig. 3B; *red circles*) and the external cadaverine level (Fig. 3C; *red circles*) did not increase beyond the values of the strain harboring CadC_T229A_E447Q. This indicates that the absolute level of *cadBA* transcription in the CadC_Y453I harboring strain is somewhat lower compared to the other strains[1].

To test whether this behavior is consistent with our model, we again fitted the parameters K_c, n_c and C_0, keeping all other parameters fixed to the values of the wild type CadC. Our fit data indicated that cadaverine sensing was indeed significantly impaired in the strain harboring CadC_Y453I (Fig. 4; *red line*), as reflected in a more than ten-fold reduction in cadaverine affinity (K_c = 2.3 mM) and a very low Hill coefficient (n_c = 0.3). Nonetheless, our model predicted a transient *cadBA* expression even for this largely cadaverine-insensitive CadC variant. Within our model the delayed down-regulation of *cadBA* expression arose from the CadBA-dependent depletion of lysine from the medium, which acts as an additional negative feedback directly on the input stimulus. Indeed, during the course of the experiment the cells converted the entire lysine (10 mM initially added) into cadaverine (approx. 10 mM after 150 min, Fig. 3C).

Beside the direct effects on cadaverine binding, also C_0, the maximal number of active $(CadC\_Y453I)_2$ dimers, was reduced by 25%, indicating that the substitution in CadC_Y453I also affected the spontaneous dimerization capability of CadC. The latter effect would be plausible as the substituted residue is located at the dimerization interface of CadC (Fig. 2). Indeed, absolute ß-galactosidase activities at inducing conditions (pH 5.8, 10 mM lysine; no cadaverine) for CadC_Y453I were reduced by 20 – 30% compared to the wild type protein (data not shown). Taken together, these results reveal that $(CadC\_Y453I)_2$ behaves like "cadaverine-blind" CadC, indicating that Tyr453 is of pivotal importance for cadaverine-sensing *in vivo*.

DISCUSSION

Adaptation of gene expression in response to external stimuli is crucial for survival under stress conditions. Here, we studied feedback inhibition of the acid stress response Cad module by cadaverine. Intriguingly, the sensor and transcriptional activator CadC simultaneously integrates three external input signals, transduces the signals into the cytoplasm and activates transcription. Thus, all signaling capabilities are concentrated in a single polypeptide. In this study we set out to investigate the cadaverine-dependent contribution to signal perception in CadC. Based on its crystal structure, the finding of an internal cavity and corresponding docking studies we postulated that cadaverine binding occurs in the cavity of the periplasmic domain (12). Site directed mutagenesis of amino acids putatively involved in cadaverine binding and subsequent investigation of the inhibition by cadaverine led to the identification of CadC variants with only moderately altered cadaverine-dependent regulation. Altogether 19 variants with amino acid substitutions in the central periplasmic cavity were tested for alterations in cadaverine-dependent inhibition of *cadBA* expression. Among those, cadaverine sensing by CadC_T229A_E447Q and CadC_T229A_Q421A_E447Q was affected

significantly, as these variants still activated *cadBA* expression in the presence of 10.4 mM cadaverine (residual activity approx. 20%). Accordingly, higher cadaverine concentrations were needed to reduce *cadBA* expression to 50% (3.6 mM and 3.8 mM cadaverine, respectively). However, combinations of up to three amino acid substitutions within the central cavity did not significantly potentiate the effect. If the internal cavity would be the major cadaverine binding site, one would expect that amino acid substitutions in this area of CadC have a more drastic effect on cadaverine-mediated repression.

Therefore, data from Dell *et al*. (27) motivated us to test different CadC variants with substitutions in the region of the patch of acidic amino acids previously identified to be crucial for detection of acidification of the external milieu (11). As polyamine binding is generally mediated through interaction with acidic and aromatic amino acids (22-26), Tyr453 in this region appeared as a promising candidate for coordination of one amine group of cadaverine. Substitution of Tyr453 against alanine, isoleucine or phenylalanine revealed that the aromatic character at this position is essential for cadaverine-dependent repression. Hence, Tyr453 might represent one binding partner. Since Leu474, Thr475 and Leu479 were also crucial for cadaverine-dependent inhibition it is speculated that these residues may contact the carbon backbone of cadaverine. Thr475 could possibly assist by binding of the second amine group as it is located within the correct range to bind a bent cadaverine molecule (approx. 5 Å). It is also conceivable, however, that there is no pronounced interaction partner for the second amine group. An inhibitory effect on *cadBA* expression has been shown for higher concentrations of agmatine, putrescine and lysine derivatives (9), indicating that variable chain lengths of the inhibitory ligands are tolerated. Altogether, amino acids around and within the patch of acidic amino acids are pivotal for cadaverine sensing, since CadC variants carrying substitutions in this area activated *cadBA* expression despite the presence of cadaverine.

To better understand the mechanisms behind cadaverine-dependent inhibition of CadC *in vivo*, we experimentally recorded the dynamics of the Cad module with two cadaverine-sensing impaired CadC variants and employed a mathematical model to extract key molecular parameters for the interaction between cadaverine and CadC. With the help of the model the affinity parameter for cadaverine was determined to be 170 µM. This is in good agreement with the *in vitro* determined affinity of $CadC_{pd}$ for cadaverine (10). It should be noted that in our *in vivo* reporter assays significantly higher cadaverine concentrations were needed to inactivate wild type CadC (Table 1). This discrepancy might be explained by the exogenous addition of cadaverine in our reporter assays, while cadaverine is endogenously produced in our dynamical experiments: Cadaverine reportedly binds to porins and reduces the outer membrane permeability. This effect is even larger when cadaverine exerts its effect from the periplasmic side (29-31). Therefore, internally produced cadaverine is captured in the periplasm as diffusion is limited through the cadaverine-inhibited porins. As externally added cadaverine is also hindered in diffusing into the periplasm it is conceivable that a much higher cadaverine concentration is needed to exert the inhibitory effect on CadC, when added from the outside.

Among all fit parameters, we found that the Hill exponents were most significantly affected by amino acid substitutions in CadC. Specifically, the Hill exponent of 3.5 determined for wild type CadC points to at least four cadaverine binding sites that might cooperatively inactivate the $CadC_2$ dimer. It is known, that the Hill exponent gives a lower limit on the number of interacting sites in positively cooperating systems (28). This finding is in agreement with a model according to which one binding site is located in the internal cavity of each monomer and a second binding site at the dimer interface of each monomer, resulting in four distinct cadaverine binding sites in a CadC dimer. Interestingly, both the substitutions in the central cavity (CadC_T229A_E447Q) as well as the substitution at the dimer interface (CadC_Y453I) drastically reduced the Hill exponent to 0.5 and 0.3, respectively. Since the Hill exponents were reduced by more than 2 in each case (as naively expected for the removal of two binding sites per dimer), these results suggested that the four sites do not modulate $CadC_2$ activity in a simple, additive manner. Such non-additive effects may result, e.g., from allosteric effects, for which binding to a single site induces small conformational changes in $CadC_2$ wild type, such

that binding to the second site is facilitated. Indeed, when binding to the central cavity was impaired (CadC_T229A_E447Q), cadaverine only weakly affected the $CadC_2$ activity (green line in Fig. 4; Hill exponent of 0.5). This effect is even stronger if binding to the dimerization site is impaired (CadC_Y453I), where cadaverine has basically no effect on the proportion of active $CadC_2$ (*red line* in Fig. 4; Hill exponent of 0.3). It is important to note that this hierarchy was also detected in our reporter assays, where CadC_Y453I was almost immune against cadaverine while CadC_T229A_E447Q still responded to cadaverine. Generally, the Hill exponents smaller than one observed in these mutants may point to a "negative cooperative" interaction between the residual binding sites, such that binding to one site impedes binding to the second site. However, heterogeneous mixtures of binding sites with different affinities can also lead to an apparent Hill exponent smaller unity (28). From our data it is not possible to discriminate between both effects. It is nevertheless intriguing that changes in single amino acids can change the cooperativity from positive to negative or to no cooperativity (32,33), indicating that the Hill exponent can be readily tuned by evolutionary forces. Thus, we expect that the precise value of the Hill exponent for the interaction between cadaverine and CadC is optimized for its function *in vivo*, for which the sensitivity and range of response have to be carefully adjusted.

Based on our results we propose the following sequential model for the inactivation of CadC by cadaverine: Each CadC monomer has a binding site in its central periplasmic cavity and a binding site at the dimerization interface (Fig. 5, *white hexagons*). In its activated conformation (low pH and high lysine; Fig. 5A) CadC forms a homodimer, which grants cadaverine access to the central cavities, but buries the binding sites at the dimerization interface. Upon binding of cadaverine to the central cavities (Fig. 5B), small conformational changes in $CadC_2$ expose the binding sites at the dimerization interface. Cadaverine binding to the site at the dimerization interface possibly acts like a wedge that leads to disruption or inhibition of the active $CadC_2$ conformation, and consequently inactivates CadC (Fig. 5C). If the binding sites within the central cavities are missing (CadC_T229A_E447Q; Fig. S4 B), the initial conformational changes are not induced. Therefore, access to the interface binding sites is sterically hindered, and inhibition of this variant occurs at higher cadaverine concentrations compared to the wild type. In contrast, the variant in which binding at the interface is impaired, binding to the internal cavities occurs at similar cadaverine levels as in the wild type (Fig. S4 C). However, the final inactivation step occurs at even higher cadaverine concentrations.

In summary, the one-component system CadC integrates the inducing (H^+) and inhibiting signals (cadaverine) at the periplasmic dimerization interface. It was already suggested that upon stimulus perception active dimer formation is promoted, e.g., through charge neutralization of repelling surfaces (11). Here we found, that the inhibitor, cadaverine, binds to the same protein area and possibly impairs active dimer formation - most likely due to spatial/conformational constraints. As vast amounts of protein are produced upon activation of *cadBA* transcription [CadA increases to up to 2% of cellular proteins (6)], the sophisticated regulatory mechanism of the Cad module seems to protect the cell against wasting energy under unfavourable conditions. For instance, the cadaverine level might serve as a proxy for the abundance of CadA and CadB, such that the negative feedback via cadaverine asserts homeostatic control of their levels. In that, the transient induction of *cadBA* supports acid stress response by a "produce-to-demand" mechanism that minimizes the cost of *cadBA* induction and saves cellular resources for the investment in other cell-protecting processes.

REFERENCES

1. Bearson, S., Bearson, B., and Foster, J. W. (1997) *FEMS Microbiol Lett* **147**, 173-180
2. Chung, H. J., Bang, W., and Drake, M. A. (2006) *Comprehensive Reviews in Food Science and Food Safety* **5**, 52-64
3. Watson, N., Dunyak, D. S., Rosey, E. L., Slonczewski, J. L., and Olson, E. R. (1992) *J Bacteriol* **174**, 530-540
4. Gale, E. F., and Epps, H. M. (1944) *Biochem J* **38**, 232-242
5. Meng, S. Y., and Bennett, G. N. (1992) *J Bacteriol* **174**, 2659-2669
6. Stim-Herndon, K. P., Flores, T. M., and Bennett, G. N. (1996) *Microbiology* **142 (Pt 5)**, 1311-1320
7. Sabo, D. L., Boeker, E. A., Byers, B., Waron, H., and Fischer, E. H. (1974) *Biochemistry* **13**, 662-670
8. Fritz, G., Koller, C., Burdack, K., Tetsch, L., Haneburger, I., Jung, K., and Gerland, U. (2009) *J Mol Biol* **393**, 272-286
9. Neely, M. N., Dell, C. L., and Olson, E. R. (1994) *J Bacteriol* **176**, 3278-3285
10. Tetsch, L., Koller, C., Haneburger, I., and Jung, K. (2008) *Molecular microbiology* **67**, 570-583
11. Haneburger, I., Eichinger, A., Skerra, A., and Jung, K. (2011) *J Biol Chem* **286**, 10681-10689
12. Eichinger, A., Haneburger, I., Koller, C., Jung, K., and Skerra, A. (2011) *Protein Sci* **20**, 656-669
13. Taylor, R. G., Walker, D. C., and McInnes, R. R. (1993) *Nucleic Acids Res* **21**, 1677-1678
14. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103-119
15. Miller, J. H. (1992) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
16. Epstein, W., and Kim, B. S. (1971) *J Bacteriol* **108**, 639-644
17. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., and Mayhew, G. F. (1997) *Science* **277**, 1453
18. Soksawatmaekhin, W., Kuraishi, A., Sakata, K., Kashiwagi, K., and Igarashi, K. (2004) *Molecular microbiology* **51**, 1401-1412
19. Quick, M., and Javitch, J. A. (2007) *Proc Natl Acad Sci U S A* **104**, 3603-3608
20. Scatchard, G. (1949) *Annals of the New York Academy of Sciences* **51**, 660-672
21. Kuper, C., and Jung, K. (2005) *J Mol Microbiol Biotechnol* **10**, 26-39
22. Kashiwagi, K., Kuraishi, A., Tomitori, H., Igarashi, A., Nishimura, K., Shirahata, A., and Igarashi, K. (2000) *Journal of Biological Chemistry* **275**, 36007
23. Kashiwagi, K., Pistocchi, R., Shibuya, S., Sugiyama, S., Morikawa, K., and Igarashi, K. (1996) *Journal of Biological Chemistry* **271**, 12205
24. Soksawatmaekhin, W., Uemura, T., Fukiwake, N., Kashiwagi, K., and Igarashi, K. (2006) *Journal of Biological Chemistry* **281**, 29213
25. Sugiyama, S., Vassylyev, D. G., Matsushima, M., Kashiwagi, K., Igarashi, K., and Morikawa, K. (1996) *Journal of Biological Chemistry* **271**, 9519
26. Vassylyev, D. G., Tomitori, H., Kashiwagi, K., Morikawa, K., and Igarashi, K. (1998) *Journal of Biological Chemistry* **273**, 17604
27. Dell, C. L., Neely, M. N., and Olson, E. R. (1994) *Mol Microbiol* **14**, 7-16
28. Abeliovich, H. (2005) *Biophysical journal* **89**, 76-79
29. Iyer, R., and Delcour, A. H. (1997) *Journal of Biological Chemistry* **272**, 18595
30. Samartzidou, H., and Delcour, A. H. (1999) *J Bacteriol* **181**, 791
31. Samartzidou, H., Mehrazin, M., Xu, Z., Benedik, M. J., and Delcour, A. H. (2003) *J Bacteriol* **185**, 13
32. Ikeda, Y., Taniguchi, N., and Noguchi, T. (2000) *J Biol Chem* **275**, 9150-9156
33. Kolodziej, A. F., Tan, T., and Koshland, D. E., Jr. (1996) *Biochemistry* **35**, 14782-14792

Acknowledgments- The authors are grateful to K. Burdack for excellent technical assistance. We thank Christoph Küper for providing strain *E. coli* MG155Δ*cadC.*

FOOTNOTES

‡ Both authors contributed equally to this work.

This work was supported by the Excellence Clusters "Nanosystems Initiative Munich", "Center for integrated Protein Science Munich (Exc114/1)" and through grants GE 1098/4-1 and JU270/5-1 by Deutsche Forschungsgemeinschaft. I. H. was supported by a research scholarship of the Elite Network of Bavaria and granted financial support by CiPS^{M} women.

1 Note that Fig. 3A only displays the mRNA levels relative to its maximal value, since those are much better determined than their absolute values.

FIGURE LEGENDS

FIGURE 1. Inhibitory effect of cadaverine on CadC-dependent *cadBA* expression. Reporter gene assays were performed with *E. coli* EP314 (*cadC1*::*Tn10, cadA'*::*lacZ*) that was complemented with plasmid-encoded CadC (wild type) or the indicated CadC variants. An overnight culture (pH 7.6) was shifted to pH 5.8 with the indicated cadaverine concentrations (0.65; 1.3; 2.6; 5.2; 10.4 mM) and harvested in the mid-logarithmic growth phase. ß-galactosidase activity was measured and the percentage of residual activity was calculated in relation to the same condition without cadaverine. Results are given as mean from at least three independent experiments. For clarity, the corresponding standard deviations were omitted. A. CadC variants with substitutions in the central cavity. B. CadC variants with substitutions close to the patch of acidic amino acids that do not influence cadaverine dependent inhibition. C. CadC variants with substitutions close to the patch of acidic amino acids that influence cadaverine dependent inhibition.

FIGURE 2. Localization of cadaverine-sensing residues in a $CadC_{pd}$ dimer. Left upper corner: simplified view of the $CadC_{pd}$ dimer. Colors and domain organization as in the large picture. White hexagons indicate cadaverine binding sites. Cartoon representation of $(CadC_{pd})_2$: The dimer interface is indicated by a dashed line. Residues involved in cadaverine binding at the interface, which were investigated in the experiments are presented as red spheres (Tyr453, Leu474, Thr475 and Leu479) or orange spheres (Asn478). Blue spheres represent the residues, that when substituted did not lead to an altered cadaverine-dependent inhibition (His240, Glu468, Asp471). Residues important for cadaverine sensing in the central cavity are depicted as black spheres (Asp225, Thr229, Glu447). The two monomers are depicted in light and dark colors, respectively (grey: N-terminal subdomain, green: C-terminal subdomain). The dimer is presented in top view towards the cytoplasmic membrane.

FIGURE 3. Induction kinetics of the Cad module *in vivo* (A-C) and *in silico* (D-E) for *E. coli* strains harboring different CadC variants (see legend). For a detailed list of all model parameters please refer to Table S2.

FIGURE 4. Simulated *in vivo* activities of wild type $CadC_2$ dimer (*black line*), $(CadC\_T229A\_E447Q)_2$ dimer (*green line*) and $(CadC\_Y453I)_2$ dimer (*red line*) as a function of external cadaverine. All curves were obtained by Eq. (2) with the corresponding parameters estimated for the *in vivo* data, see Table S2.

FIGURE 5. Model for cadaverine-dependent regulation of *cadBA* expression. CadC dimers of the periplasmic domain are presented schematically as the two monomers (dark and light colors) in top view. The monomers consist of two subdomains represented by a grey and green cylinder (cf. Fig. 2). Cadaverine binding sites are indicated as hexagons. Cadaverine is depicted as molecule with five carbon atoms and two amine groups (spheres). A. In the presence of both stimuli (low pH and high lysine) CadC is **active** (as a dimer) and induces *cadBA* transcription. B. As soon as cadaverine reaches the threshold level it binds to the binding site within the internal cavity and evokes (conformational) changes that render the binding sites at the interface accessible. Still, CadC is **active**. C. After cadaverine binds to the interface binding site, CadC is **inactive**; probably by disturbance of the dimer conformation. Lysine regulation accomplished by LysP was omitted for simplicity.

TABLE 1. Effect of amino acid substitutions within the central cavity and at the dimerization interface on cadaverine-dependent inhibition of *cadBA* expression.

Figure 1

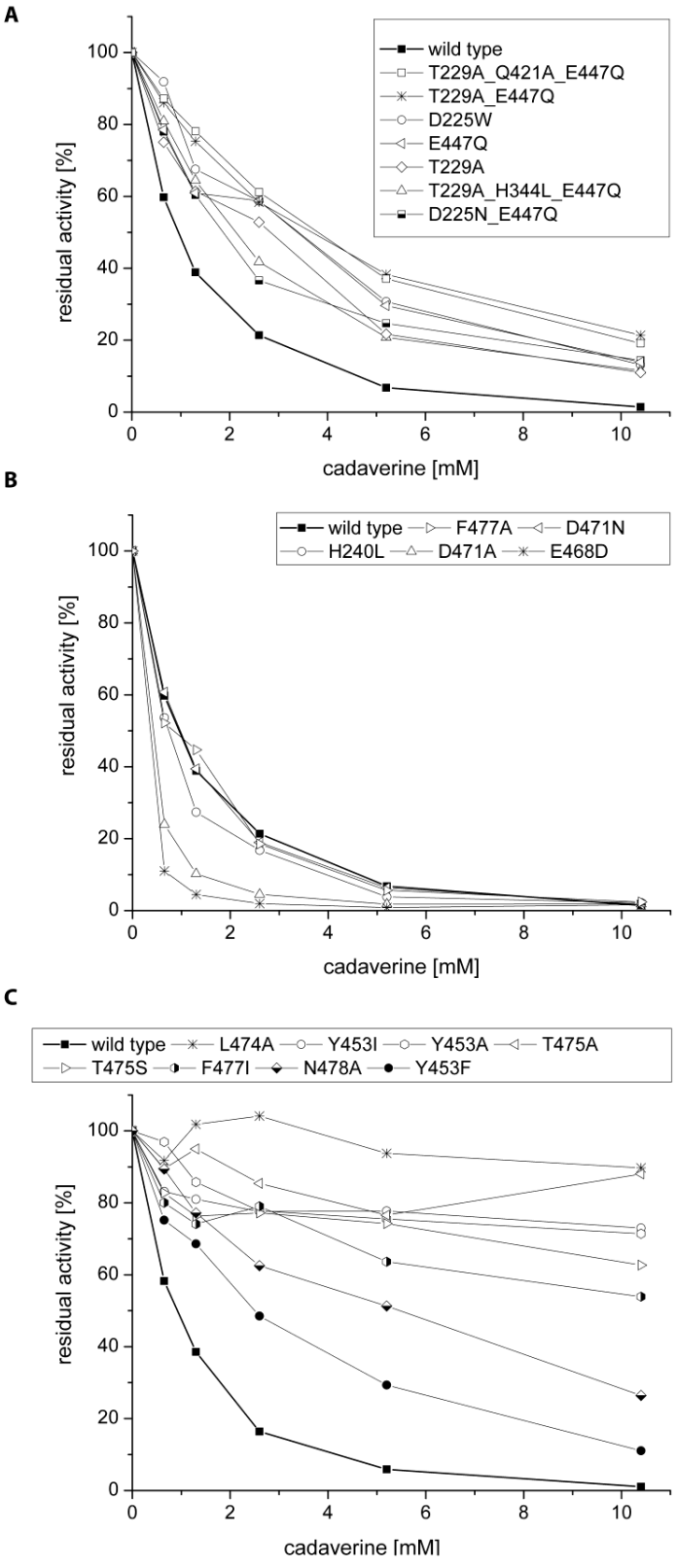

A
wild type
T229A_Q421A_E447Q
T229A_E447Q
D225W
E447Q
T229A
T229A_H344L_E447Q
D225N_E447Q
residual activity [%]
cadaverine [mM]
B
wild type
F477A
D471N
H240L
D471A
E468D
residual activity [%]
cadaverine [mM]
C
wild type
L474A
Y453I
Y453A
T475A
T475S
F477I
N478A
Y453F
residual activity [%]
cadaverine [mM]

Figure 2

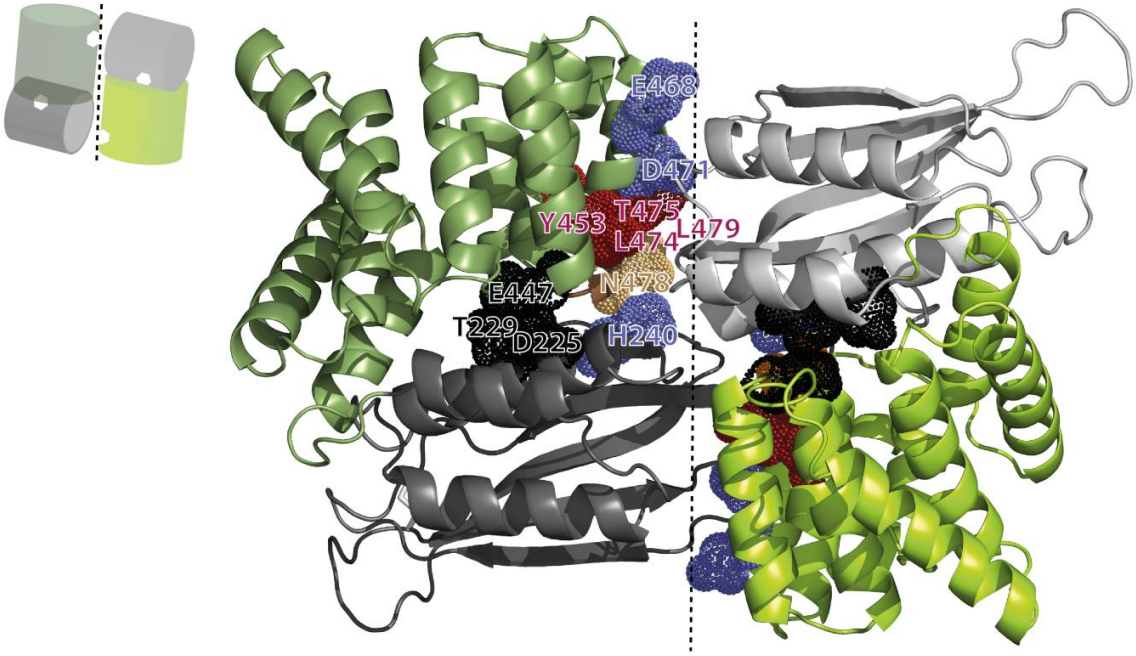


Figure 3

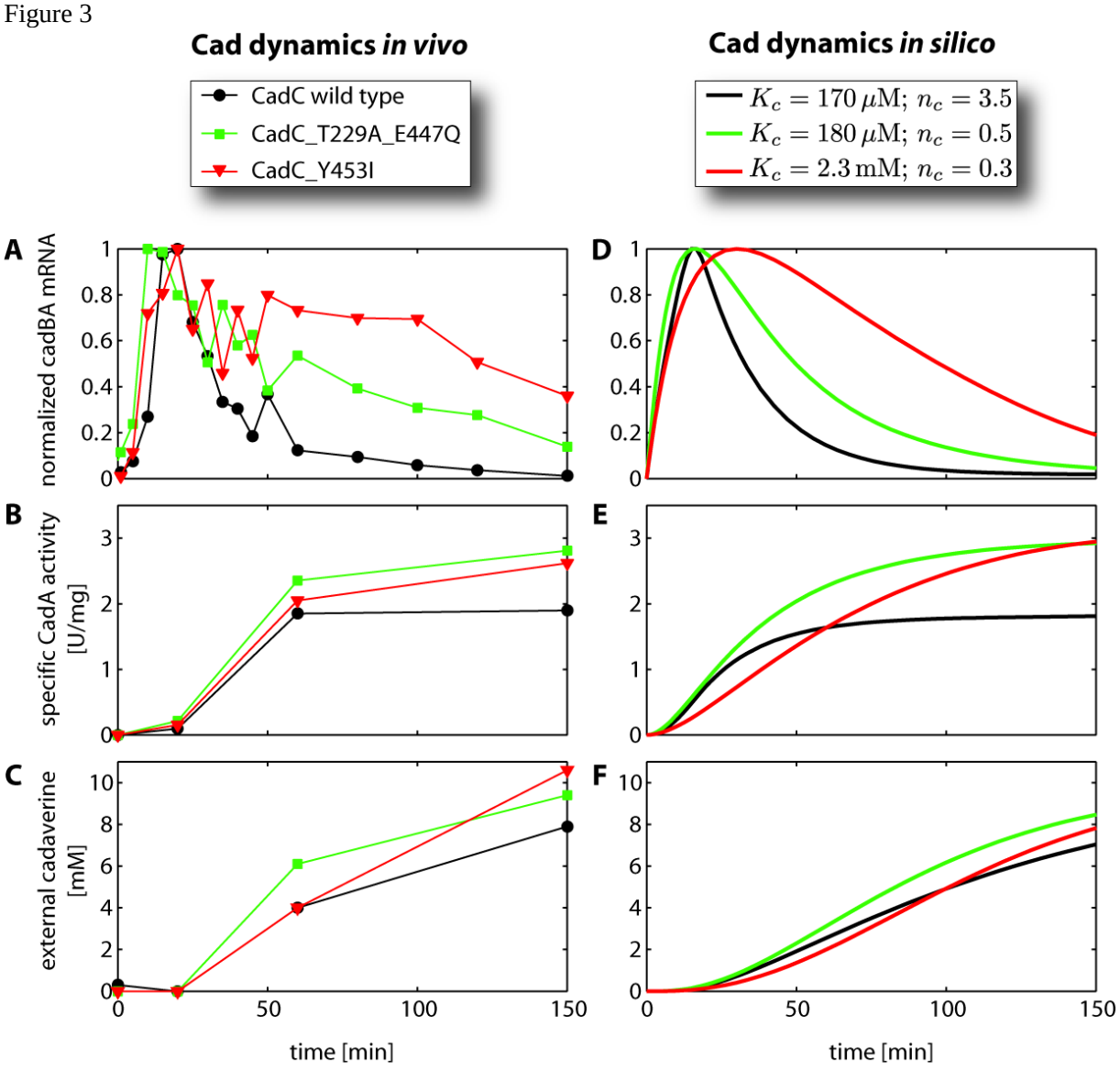

Cad dynamics *in vivo*
CadC wild type
CadC_T229A_E447Q
CadC_Y453I
Cad dynamics *in silico*
$K_c = 170\,\mu\mathrm{M};\ n_c = 3.5$
$K_c = 180\,\mu\mathrm{M};\ n_c = 0.5$
$K_c = 2.3\,\mathrm{mM};\ n_c = 0.3$
A
normalized cadBA mRNA
B
specific CadA activity [U/mg]
C
external cadaverine [mM]
time [min]
D
E
F
time [min]

Figure 4

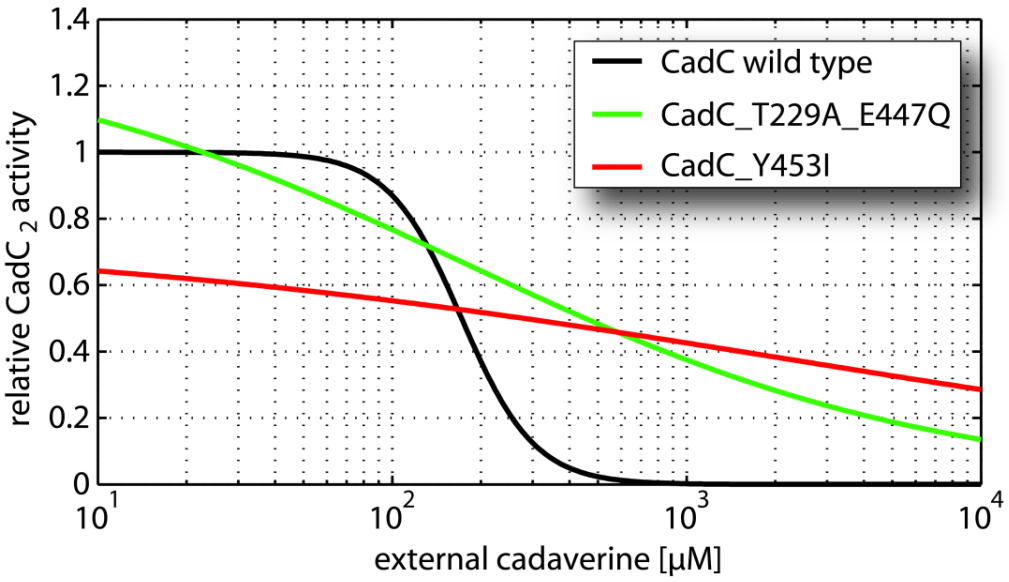


Figure 5

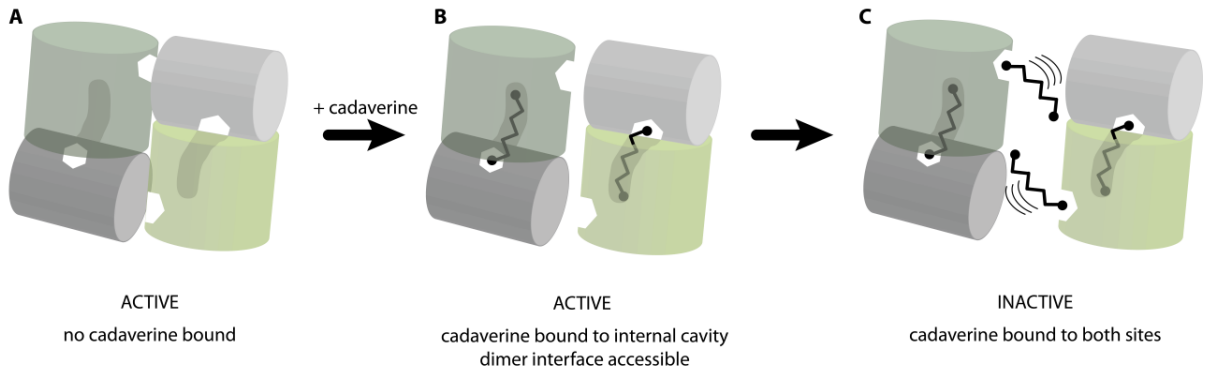


Table 1

central cavity			dimerization interface		
CadC variant	**IC_{50} (50% inhibition [mM] cadaverine)**	**residual activity at 10.4 mM cadaverine [%]**	**CadC variant**	**IC_{50} (50% inhibition [mM] cadaverine)**	**residual activity at 10.4 mM cadaverine [%]**
Y374A	0.6	8.6 ± 8.0	E468D	0.4	1.6 ± 0.3
Y374F	0.8	1.7 ± 0.4	D471A	0.4	1.8 ± 0.5
D225A	0.9	1.7 ± 1.3	H240L	0.8	2.2 ± 2.1
W450A	0.9	7.0 ± 10.5	D471N	1.0	1.7 ± 0.4
wild type	**1.0**	1.4 ± 1.7	**wild type**	**1.0**	1.1 ± 0.9
W450F	1.1	6.9 ± 6.4	F477A	1.4	1.9 ± 1.5
D225N	1.2	7.0 ± 8.9	Y453F	2.5	11.1 ± 1.3
D225N_E447Q	1.8	14.4 ± 12.2	N478A	5.5	26.4 ± 1.9
Q421A	1.8	4.1 ± 0.5	F477I	>> 10	53.9 ± 6.6
H344L_E447Q	1.9	1.3 ± 1.3	L474A	>> 10	89.7 ± 8.3
T229A_H344L	2.0	9.2 ± 4.0	Y453A	>> 10	71.4 ± 9.4
H344L	2.0	7.0 ± 1.6	Y453I	>> 10	73 ± 11.9
D225N_T229A_E447Q	2.1	7.0 ± 2.0	T475A	>> 10	88.0 ± 14.0
T229A_H344L_E447Q	2.2	11.6 ± 3.8	T475S	>> 10	62.7 ± 3.5
T229A	2.8	11.0 ± 4.7			
E447Q	3.4	13.9 ± 6.9			
D225W	3.4	13.2 ± 2.0			
T229A_E447Q	3.6	21.4 ± 1.3			
T229A_Q421A_E447Q	3.8	19.1 ± 4.2			

Supplemental Data

The feedback-inhibitor cadaverine suppresses the pH response by binding to the pH susceptible site of CadC in *Escherichia coli*

Ina Haneburger‡, Georg Fritz‡, Nicole Jurkschat, Larissa Tetsch, Andreas Eichinger, Arne Skerra, Ulrich Gerland and Kirsten Jung

‡ Both authors contributed equally to this work

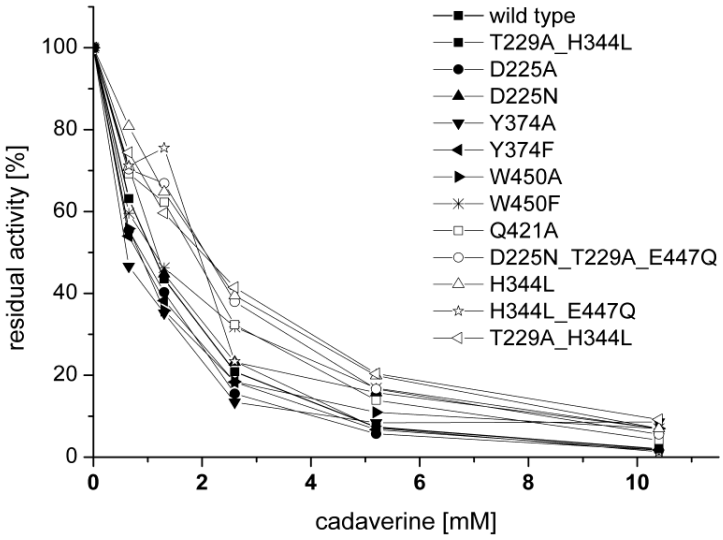


FIGURE S1. Inhibitory effect of cadaverine on *cadBA expression.* Reporter gene assays were performed with *E. coli* EP314 (*cadC1*::*Tn10, cadA'*::*lacZ*) that was complemented with plasmid-encoded CadC (wild type) or the indicated CadC variants. An overnight culture (pH 7.6) was shifted to pH 5.8 with the indicated cadaverine concentrations (0.65; 1.3; 2.6; 5.2; 10.4 mM) and harvested in the mid-logarithmic growth phase. ß-galactosidase activity was measured and the percentage of residual activity was calculated in relation to the same condition without cadaverine. Results are given as mean from at least three independent experiments. For clarity, the corresponding standard deviations are omitted.

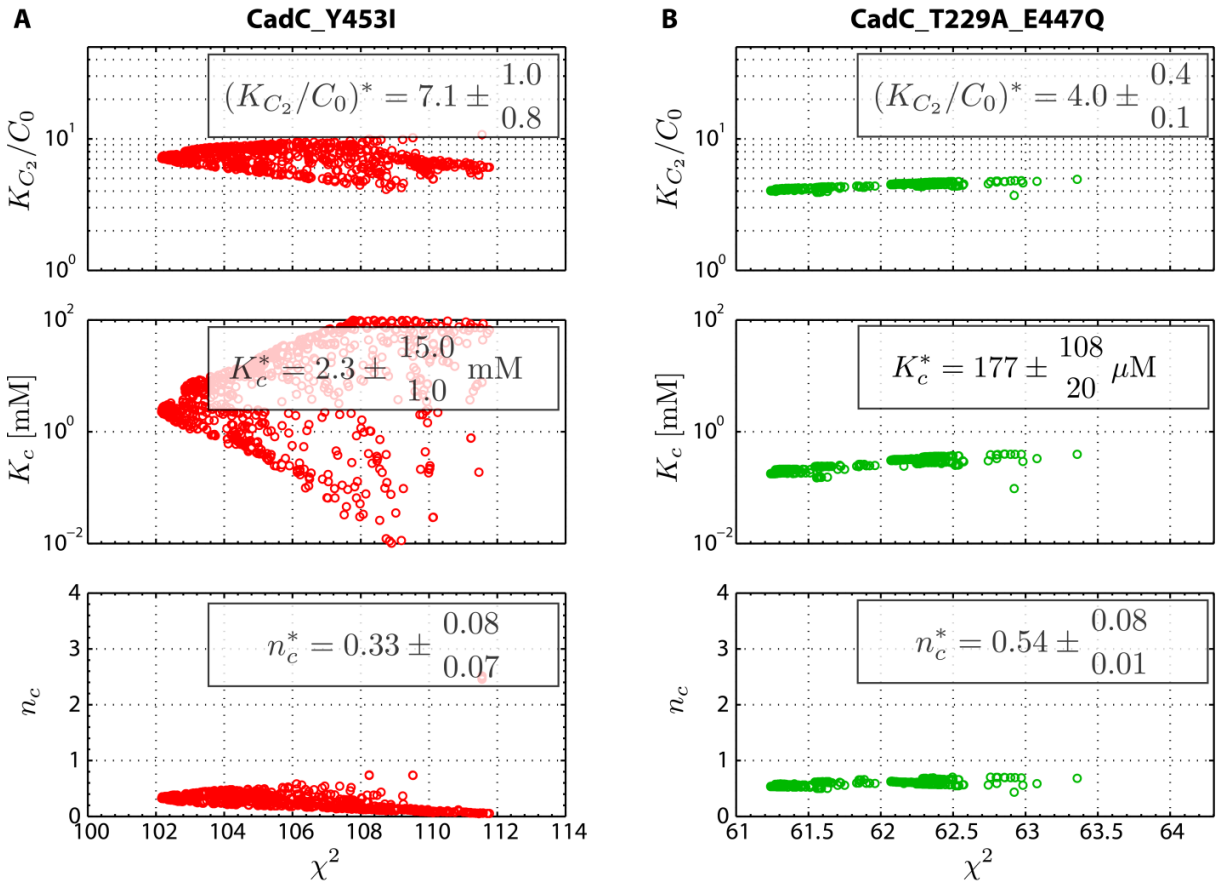


FIGURE S2. Correlations between the goodness of fit and the estimated parameters for (A) the strain harboring CadC_Y453I and (B) the strain harboring CadC_T229A_E447Q. The points correspond to local optima in the parameter space, for which the difference between the quantitative model and the experimental data in Fig. 3 of the main text is minimized (see Materials and Methods). As the fit quality increases (lower χ^2), most parameters are confined to narrow intervals, indicating that their values are well constrained by the experimental data. However, some parameters display significant variation even for the lowest χ^2 values, and we find from parameter-parameter correlation analysis in Fig. S3, that only combinations of those are well confined by our data. In all cases the y-axis corresponds to the full range of allowed parameter values.

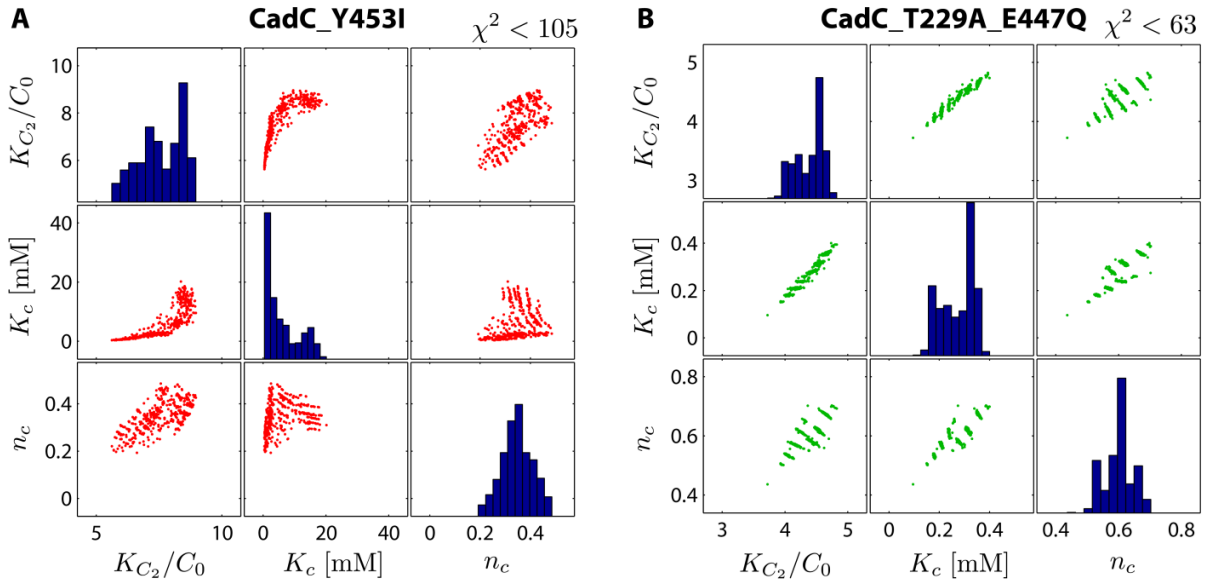


FIGURE S3. Parameter-parameter correlations between the estimated parameters for (A) strain harboring CadC_Y453I and (B) strain harboring CadC_T229A_E447Q.

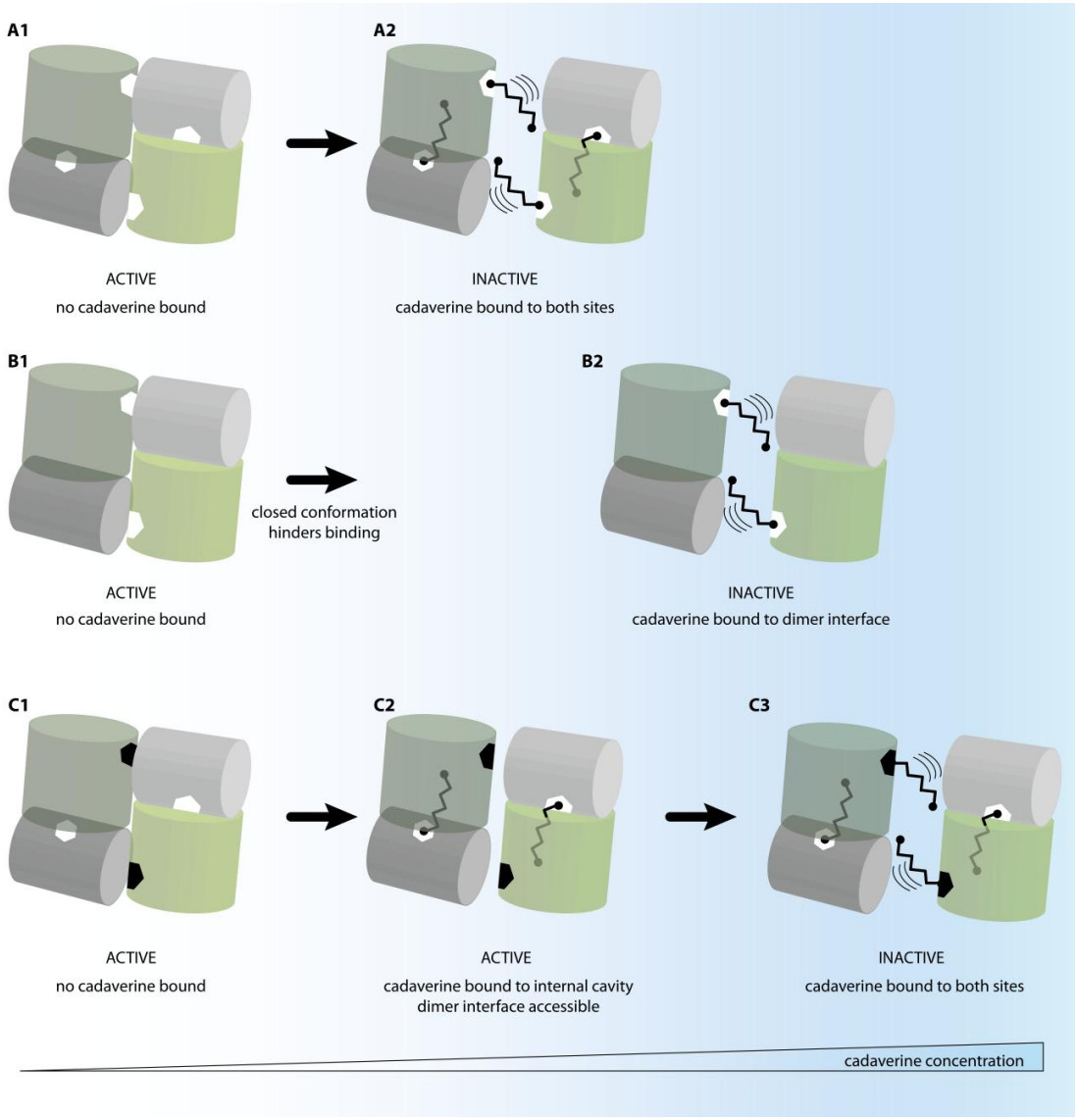


FIGURE S4. Model for cadaverine-dependent regulation of *cadBA* expression. An Increase in cadaverine concentration is indicated by blue shading. CadC dimers of the periplasmic domain are represented schematically as the two monomers (light and dark colors) in top view. The monomers consist of two subdomains represented by a grey and a green cylinder (cf. Fig. 2). Cadaverine binding sites are indicated as hexagons (white: wild type, black: binding impaired due to substitution). Cadaverine is depicted as molecule with five carbon atoms and two amine groups (spheres). A. Inactivation cycle as proposed for the wild type (cf. Fig. 5). B. Effect of substitutions in the central cavity (CadC_T229A_E447Q). Cadaverine binding to the internal cavity is significantly reduced. Therefore, initial conformational changes that would lead to exposure of the binding sites at the dimerization interface are not induced. Nonetheless they may occur at a lower rate simply due to the flexibility of proteins. In consequence, cadaverine-binding and -dependent inhibition occur at higher cadaverine concentration (B2). C. Substitutions that impair binding of cadaverine to the sites at the dimer interface (indicated by the black hexagons at the interface) would lead to a significant decrease in cadaverine-binding, in turn (nearly) prevent CadC inactivation or shift it to even higher cadaverine concentrations (C3). The initial binding step to the binding sites within the central cavities would occur at the same concentration as in the wild type (C2).

TABLE S1. Plasmids and oligonucleotides used in this study. If not indicated pET16b-cadC7 (1) was used as template.

Plasmid	Oligonucleotide	5' - 3' Sequence	Source/Reference
pET16b			Novagen
pET16b-cadC7			(1)
pET16b-cadC_D225A	CadC_D225A_sense CadC_D225A_anti	GGCGTGGGTGCTTTGGTGGCGACATC GATGTCGCCACCAAAGCACCCACGCC	This work
pET16b-cadC_D225N	CadC_D225N_sense CadC_D225N_anti	GGCGTGGGTAATTTGGTGGCGACATC GATGTCGCCACCAAATTACCCACGCC	This work
pET16b-cadC_D225W	CadC_D225W_sense CadC_D225W_as	ATAGGCGTGGGTTGGTTGGTGGCGAC GTCGCCACCAACCAACCCACGCCTAT	This work
pET16b-cadC_T229A	CadC_T229A_sense CadC_T229A_as	TTTGGTGGCGGCATCACTTAAC GTTAAGTGATGCCGCCACCAAA	This work
pET16b-cadC_H344L			(1)
pET16b-cadC_Y374A	CadC_Y374A_sense CadC_Y374A_anti	CCAGAATTTACCGCCGCGAGAGCAG This work CTGCTCTCGCGGCGGTAAATTCTGG	
pET16b-cadC_Y374F	CadC_Y374F_sense CadC_Y374F_anti	CCAGAATTTACCTTCGCGAGAGCAG CTGCTCTCGCGAAGGTAAATTCTGG	This work
pET16b-cadC_Q421A	CadC_Q421A_sense CadC_Q421A_as	CATTATATATGCAATAAAAGCGG CCGCTTTTATTGCATCACTTAAC	This work
pET16b-cadC_E447Q			(1)
pET16b-cadC_W450A	CadC_W450A_anti	CGGTGAGATATGCATCAGCTGCTTC CCGGTTCATCCCCTTCATTTCATAAA CCTTGCCAAGCAACACATAATTTAGC GCGGACATTTCAAGATC	This work
pET16b-cadC_W450F	CadC_W450F_anti	CGGTGAGATATGCATCAGCTGCTTC CCGGTTCATCCCCTTCATTTCATAAA VVTTGCCAAGCAACACATAATTTAGG AAGGACATTTCAAGATC	This work
pET16b-cadC_T229A_E447Q	combination of pET16b-cadC_T229A and pET16b-cadC_E447Q by NcoI and SacI digestion		This work
pET16b-cadC_H344L_E447Q	PCR with with oligonucleotide CadC_H334L (1) and pET16b-cadC_E447Q as template		This work
pET16b-cadC_T229A_H344L	combination of pET16b-cadC_T229A and pET16b-cadC_H344L by BamHI and SacI digestion		This work
pET16b-cadC_D225N_E447Q	PCR with oligonucleotides CadC_D225N_sense and CadC_D225N_anti with pET16b-cadC_E447Q as template		This work
pET16b-cadC_T229A_Q421A_E447Q	PCR with oligonucleotides CadC_D225N_sense and CadC_Q421A_as and pET16b-cadC_T229A_E447Q as template		This work
pET16b-cadC_D225N_T229A_E447Q	PCR with oligonucleotides CadC_D225N_sense and CadC_Q421A_as and pET16b-cadC_T229A_E447Q as template		This work
pET16b-cadC_T229A_H344L_E447Q	combination of pET16b-cadC_T229A and pET16b-cadC_H344L_E447Q by BamHI and SacI digestion		This work
pET16b-cadC_Y453A	CadC_Y453A	GCATTGATCTCGAGATGTCCTGGCTAAA TGCTGTGTTGC	This work
pET16b-cadC_Y453I			(1)
pET16b-cadC_Y453F	CadC_Y453F	GCATTGATCTCGAGATGTCCTGGCTAAA TTTTGTGTTGC	This work
pET16b-cadC_T475A	CadC_T475A_sense CadC_T475A_anti	CATATCTCGCCGCCTTTAATTTACGCCCC GGGGCAAAC GTTTGCCCCGGGGCGTAAATTAAAGGCG GCGAGATATG	This work
pET16b-cadC_T475S	CadC_T475S	GTGTTTGCCCCGGGGCGTAAATTAAAGG CGGAGAGATATGC	This work
pET16b-cadC_H240L			(1)
pET16b-cadC_E468D			(1)
pET16b-cadC_D471A			(1)
pET16b-cadC_D471N			(1)
pET16b-cadC_L474A			(1)
pET16b-cadC_N478A			(1)
pET16b-cadC_F477A			(1)
pET16b-cadC_F477I			(1)
pET32a			Novagen
pET32a-cadC188–512	his 6 -cadC 188–512 in pET32a		(2)
pET32a-cadC188–512_Y453I	his 6 -cadC 188–512_Y453I in pET32a		This work
pET32a-cadC188–512_T475A	his 6 -cadC 188–512_T475A in pET32a		This work
pET32a-cadC188–512_T229A_E447Q	his 6 -cadC 188–512_T229A_E447Q in pET32a		This work

TABLE S2. Parameters of the quantitative model for the wild type Cad module. LB indicates lower bound; UB, upper bound. The estimated parameter values are shown as (best-fit value) $^{+\sigma^+}_{-\sigma^-}$, where σ^+ and σ^- indicate the asymmetric standard errors in the positive direction and in the negative direction, respectively, see (3) for details.

	Parameter	LB	UB	Estimated Value
	Sensory module			
K_I	Threshold for CadC activation by lysine	1	20	$4.8 \pm {}^{3.4}_{0.7}$ mM
K_c	Threshold for CadC inactivation by cadaverine	50	1000	$172 \pm {}^{18}_{20}$ μM
n_I	Hill coefficient for CadC regulation by lysine	1	5	$1.6 \pm {}^{0.1}_{0.2}$
n_c	Hill coefficient for CadC regulation by cadaverine	1	5	$3.5 \pm {}^{0.6}_{0.6}$
pH_0	pH threshold for CadC activation	-	-	6.2
ΔpH	width of the transition from active to inactive CadC	-	-	0.5
	Expression module			
C_0/K_C	Total CadC per cell in relation to the threshold for CadC-promoter binding	0.1	10	$0.19 \pm {}^{1.20}_{0.06}$
ν_m	basal transcription rate	0.001	0.1	$2.4 \pm {}^{2.7}_{1.7} \times 10^{-2}$ min^{-1}
f	fold-change between basal and maximal transcription rate	10	1000	$80 \pm {}^{89}_{42}$
ν_p	effective translation rate	10^{-4}	10^{-1}	$7 \pm {}^{62}_{3} \times 10^{-4}$ U/min
τ_m	mRNA half-life ($= \ln 2/\lambda_m$)	1	50	$14.2 \pm {}^{1.6}_{0.7}$ min
τ_p	protein half-life ($= \ln 2/\lambda_p$)	1	10^4	$35 \pm {}^{118}_{10}$ h
v_{max}	maximal rate for lysine turnover via CadA and CadB	10^{-4}	10	$1.1 \pm {}^{1.5}_{0.3} \times 10^{-3}$ min^{-1}
K_m	effective Michaelis constant for lysine turnover via CadA and CadB	1	100	$20 \pm {}^{38}_{7}$ mM

REFERENCES

1. Haneburger, I., Eichinger, A., Skerra, A., and Jung, K. (2011) *J Biol Chem* **286**, 10681-10689
2. Tetsch, L., Koller, C., Haneburger, I., and Jung, K. (2008) *Molecular microbiology* **67**, 570-583
3. Fritz, G., Koller, C., Burdack, K., Tetsch, L., Haneburger, I., Jung, K., and Gerland, U. (2009) *J Mol Biol* **393**, 272-286

Timing and Dynamics of Single Cell Gene Expression in the Arabinose Utilization System

Judith A. Megerle,* Georg Fritz,*† Ulrich Gerland,† Kirsten Jung,‡ and Joachim O. Rädler*
*Department für Physik und CeNS, Ludwig-Maximilians-Universität, Munich, Germany; †Institut für Theoretische Physik, Universität zu Köln, Cologne, Germany; and ‡Department Biologie I, Bereich Mikrobiologie, Ludwig-Maximilians-Universität, Munich, Germany

ABSTRACT The arabinose utilization system of *Escherichia coli* displays a stochastic all-or-nothing response at intermediate levels of arabinose, where the population divides into a fraction catabolizing the sugar at a high rate (on-state) and a fraction not utilizing arabinose (off-state). Here we study this decision process in individual cells, focusing on the dynamics of the transition from the off- to the on-state. Using quantitative time-lapse microscopy, we determine the time delay between inducer addition and fluorescence onset of a GFP reporter. Through independent characterization of the GFP maturation process, we can separate the lag time caused by the reporter from the intrinsic activation time of the arabinose system. The resulting distribution of intrinsic time delays scales inversely with the external arabinose concentration, and is compatible with a simple stochastic model for arabinose uptake. Our findings support the idea that the heterogeneous timing of gene induction is causally related to a broad distribution of uptake proteins at the time of sugar addition.

INTRODUCTION

Bacteria have sophisticated signal transduction and gene regulatory networks for rapid adaptation to environmental changes. In recent years it became increasingly recognized, that the dynamical response of these biochemical reaction networks is subject to significant stochastic fluctuations (1), which can lead to heterogeneous behavior across cellular populations. Examples include the transient differentiation of *Bacillus subtilis* in its late exponential phase (2,3), bacterial persistence in *Escherichia coli* (4), and the mating pheromone response pathway in yeast (5). In many of these systems, positive feedback plays a fundamental role, since it gives rise to bistability and thereby causes two clearly distinct gene expression states (6). It has been demonstrated that biochemical noise induces stochastic transitions between the two stable states, and it was suggested that the resulting population heterogeneity provides selective advantages for colony growth in fluctuating environments (7,8).

A prototypic class of positive feedback systems are the inducible sugar utilization systems, in which bistability is caused by the autocatalytic positive feedback of the sugar on its own uptake proteins. These systems allow bacteria to grow on less favorable carbon sources than glucose: For instance, in a medium where lactose is the only energy source, *E. coli*'s lactose utilization (*lac*) system either imports and catabolizes lactose at a high rate (on-state), or it does not use lactose at all (off-state) (9). This bistable behavior has drastic effects on the behavior at the population level. When a high amount of external lactose was added to a previously uninduced culture, all cells in the population switched from the off- to the on-state. However, at lower sugar concentrations only a fraction of cells switched to the on-state while others remained in the off-state (9,10).

Here, we are interested in the dynamics of such a switching process on the single cell level. We study these dynamics in the context of the arabinose utilization (*ara*) system of *E. coli* (11), another well-characterized bistable system (see Fig. 1). In this case, arabinose is imported by the high-affinity low-capacity transporter AraFGH and the low-affinity high-capacity transporter AraE. If internal arabinose exceeds a threshold concentration, it activates AraC, which in turn promotes expression of *araFGH*, *araE*, and the genes for arabinose catabolism, *araBAD*. Siegele and Hu (12) analyzed population distributions at intermediate sugar levels, and revealed that the *ara* system displays an all-or-nothing expression pattern similar to the *lac* system. They conjectured that in uninduced cells the stochastic background expression of the *ara* regulon leads to a wide distribution of *ara* uptake proteins. Addition of arabinose would then lead to different rates of arabinose accumulation, causing heterogeneous timing of gene induction within the population. At a given time there would be a fraction of induced and a fraction of uninduced cells, and the depletion of arabinose by the metabolism of the induced cells could explain the fixation of the all-or-nothing response. This conjecture is consistent with a computational study of autocatalytic expression systems (13) and experiments which placed *araE* under the control of a constitutive promoter, finding homogeneous gene expression in the population (14–16). The dynamics of switching processes has also been studied using flow cytometry techniques, which yield a time series of population distributions of gene expression levels (17,18).

In this study, we take a different experimental approach: Rather than recording population distributions, we use quantitative time-lapse fluorescence microscopy to follow

Submitted December 5, 2007, and accepted for publication April 23, 2008.

Judith A. Megerle and Georg Fritz contributed equally to this work.

Address reprint requests to Judith A. Megerle, Tel.: 49-89-21-80-27-04; E-mail: judith.megerle@physik.uni-muenchen.de.

Editor: Herbert Levine.

0006-3495/08/08/2103/13 $2.00

doi: 10.1529/biophysj.107.127191

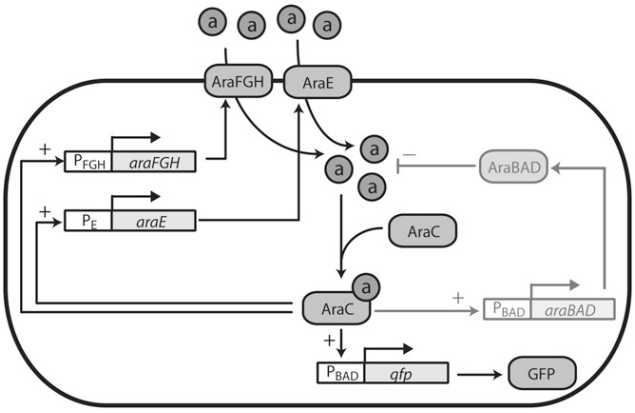


FIGURE 1 Regulatory network of the native arabinose utilization system (11), including the *gfp*-reporter module used in this study. The system consists of genes for arabinose uptake (*araE*, *araFGH*), genes for arabinose metabolism (*araBAD*), and the regulator AraC. High amounts of intracellular arabinose activate AraC, which stimulates expression from the promoters P_{BAD}, P_E, and P_{FGH}. In the absence of arabinose, AraC represses expression from P_{BAD} (not depicted). Note that we also omitted the less pronounced negative autoregulation of AraC in the absence and presence of arabinose (43), since this feedback mainly seems to provide a constant transcription factor abundance (52). As indicated in light shading, in the mutant used in this study the chromosomal *araBAD* operon is deleted, and hence the additional negative feedback on internal arabinose is avoided. As a reporter for the expression of the ara system we used a plasmid-borne *gfp* variant under the control of the P_{BAD} promoter (see Materials and Methods for details).

the expression dynamics of the switching process in many cells, individually. In a physics analogy, this is akin to following the trajectories of many particles, instead of recording their spatial density distribution at different time points. Clearly, the distributions can be obtained from the trajectories, but not vice versa; i.e., the trajectories contain more information. In this case, this additional information is particularly useful to disentangle different variables that affect the response of individual cells: The observed time-dependent fluorescence level is the final output of a series of biochemical processes, which can be grouped into two connected subsystems—an uptake module and a reporter module. Both modules experience noise, which, to a first approximation, can be subsumed into a single parameter for each module. As we will see, one can extract these two parameters for each cell by fitting an appropriate model to the fluorescence trajectory of the cell. As a result, we can directly obtain the separate distributions for these two parameters, and even measure their correlations. Note that this analysis would not have been possible based on population distributions of gene expression levels.

Using this approach, we address the question raised by Siegele and Hu (12), i.e., is the all-or-nothing response of the *ara* system associated with heterogeneous timing of gene induction, and, if so, is the heterogeneous timing causally related to a wide distribution of *ara* uptake proteins? At subsaturating sugar levels, we observe a significant delay between addition of inducer and increase of fluorescence, which is indeed broadly distributed. To clarify the origin of this delay and its broad distribution, it is necessary to separate the intrinsic lag of the GFP expression dynamics from the time-lag inherent to the stochastic arabinose uptake. To this end, we leverage our microfluidic setup to separately measure the distribution of GFP maturation times across an *E. coli* population. We also record the cell-to-cell variation of the growth rates. Using a simple quantitative model for the expression dynamics, we then extract the intrinsic timing statistics for gene induction. We find that this distribution is well described by an analytical delay time distribution derived from a stochastic model for the uptake module. Our results support the conclusion that the heterogeneous timing is indeed due to a wide distribution of *ara* uptake proteins across the population.

MATERIALS AND METHODS

Bacterial strain and plasmid

E. coli strain LMG194 (*F⁻ lacX74 galE galK thi rpsL phoA* (*Pvu*II) *ara714 leu*::*Tn10*) (19) was transformed with plasmid pBAD24-GFP (this work) using a standard method as described elsewhere (20). The gene *gfpmut3* (21) encoding the green fluorescent protein GFPmut3 was amplified by PCR with primers GFP-*Kpn*I sense (5′-TACCATGGTACCAAGTAAAGGAGAA-GAACTTTTC-3′) and GFP-*Hin*dIII antisense (5′-CATAGTAAGCTTTT-ATTTGTATAGTTCATCCATGCC-3′) using plasmid pJBA29 (22) as a template. The DNA-fragment was cut with restriction endonucleases *Kpn*I and *Hin*dIII, and was then ligated into similar treated vector pBAD24 (19), resulting in plasmid pBAD24-GFP. The correct insertion of the fragment was verified by restriction analysis as well as by DNA sequence analysis.

Growth conditions

Cells were grown in LB medium (23) or M63 minimal medium (19) containing 0.2% (w/v) glycerol as C-source. When indicated, 0.01%, 0.02%, 0.05%, or 0.2% (w/v) arabinose was added to induce GFP expression. Bacteria were inoculated from single colonies grown on LB agar plates and grown overnight (37°C, shaking at 300 rpm) in M63 medium. Overnight cultures were diluted 1:50 into fresh M63 medium and cultured for 2 h. Bacteria were subsequently diluted in prewarmed medium to an appropriate density and were then applied to one channel of a poly-L-lysine-coated microfluidic chamber (μ-Slide VI; Ibidi, Martinsried, Germany). The slide was then incubated at 37°C for several minutes. By softly flushing the channel with prewarmed medium supplemented with the desired arabinose concentration, gene expression was induced and the sample was rinsed at the same time. After the preparation procedure, the vast majority of the bacteria adhered with their long axis parallel to the surface.

Time-lapse microscopy

Time-lapse experiments were performed on a fully automated inverted microscope (Axiovert 200M, Zeiss, Oberkochen, Germany) equipped with a motorized stage (Prior Scientific, Cambridge, UK). All devices were controlled by Andor IQ software (Andor, Belfast, Northern Ireland). Fluorescence illumination was provided by an X-cite120 light source (EXFO, Quebec, Canada). An appropriate filter set (excitation: 470/40; beamsplitter 495; emission: 525/50; filter set Nr 38; Zeiss) was used. Bright field and fluorescence images of several fields in one sample were acquired every

5 min with a highly sensitive EMCCD camera (iXon DV885; Andor) through an oil-immersion 100× plan-neofluar objective with NA 1.3 (Zeiss), with acquisition times of 0.1 s to 0.2 s. To further prevent photobleaching and photodamage all light sources were shuttered between exposures and an orange filter was used in the bright-field light path. The temperature in the sample environment was maintained at 37°C using a custom-built heating box. Focalcheck fluorescence microspheres (Invitrogen, Karlsruhe, Germany) were used to correct for output variations of the lamp.

Data analysis

ImageJ (24) and Igor Pro 4.0 (WaveMetrics, Lake Oswego, OR) were used for data analysis. Cell outlines were created by thresholding the bright-field images. Total fluorescence was measured as the sum over all pixel values within the outline in the corresponding background-corrected fluorescence image. Time traces were assembled by tracking the cells manually. As photobleaching was found to be negligible for the given experimental system, fluorescence traces were fitted without further processing.

Measurement of the GFP maturation time distribution in vivo

The maturation time in single cells was determined using an approach similar to the one established in Gordon et al. (25): Translation was blocked by the addition of 200 μg/ml chloramphenicol, 30 min after the induction of *gfp*-expression with 0.2% arabinose. Fluorescence images were acquired every 3–5 min before and after inhibition. As this measurement was more sensitive, the illumination was reduced and the EM gain of the camera was used. Photobleaching could thus again be neglected. Cellular fluorescence was determined by summing all pixel values above the background level for each bacterium. This method is qualitatively equal to the use of cell outlines as described above, but can only be applied if the range of fluorescence values is limited and bacteria do not grow strongly. The resulting maturation time courses were fitted by an exponential function.

RESULTS

Single cell induction kinetics

To study the induction kinetics of the *ara* system, we use an *E. coli* strain where both *araBAD* and *araC* are deleted (19). It is transformed with the reporter plasmid pBAD24-GFP, containing the *araC* gene and the rapidly maturing GFP variant *gfpmut3* (21) (which is under the control of the P_{BAD} promoter; see Materials and Methods). The *araC* gene is supplied on the plasmid to guarantee full functionality of the DNA loop required for repression of P_{BAD} in the absence of arabinose (11) and to provide the proper stoichiometry of transcription factors and P_{BAD} promoters. The chromosomal deletion of *araBAD* avoids the negative feedback of the internal arabinose catabolism. This feedback complicates the system, but is irrelevant for our questions, which focus on the kinetics of the induction when arabinose first becomes available externally. The gene regulatory circuit of our system is illustrated in Fig. 1.

To perform the time-lapse fluorescence microscopy, we introduce the bacteria into a microfluidic chamber, where they attach to the poly-L-lysine coated chamber wall. The microfluidic chamber provides homogeneous external conditions for the bacteria and can be used to rapidly exchange the medium. At $t = 0$ min, we induce the bacteria with 0.2% (13.3 mM), 0.05% (3.33 mM), 0.02% (1.33 mM), or 0.01% (0.66 mM) arabinose, and then record the time-evolution of GFP fluorescence in single cells. Representative fluorescence trajectories for the highest (0.2%) and the lowest (0.01%) arabinose concentration are shown in Fig. 2, *a* and *b*, respectively.

For all arabinose concentrations, the individual time-traces of each cell appear rather smooth and deterministic, whereas there is a significant variation in the response from cell to cell. We also observe a time lag between the addition of arabinose and the onset of fluorescence. With decreasing arabinose concentration, the typical lag time becomes longer, and its cell-to-cell variation becomes more pronounced. Below, we

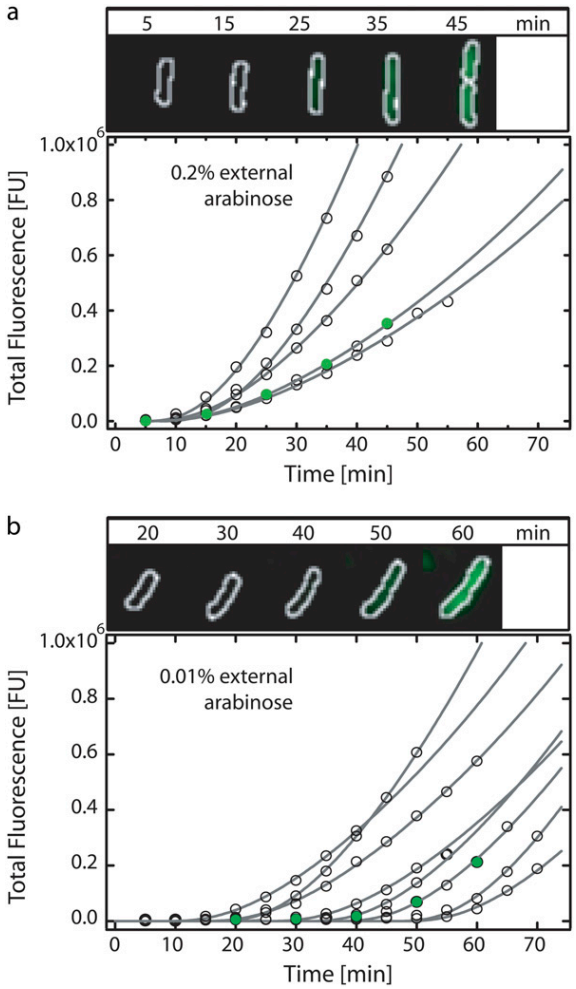


FIGURE 2 Examples of single cell induction kinetics of the arabinose utilization network. Cells were induced at $t = 0$ min with 0.2% arabinose (*a*) and 0.01% arabinose (*b*) (*open circles*). The traces were analyzed up to the first cell division, which results in different numbers of data points in the traces. Fits of the deterministic gene expression function in Eq. 1 to the data are shown as solid lines. The image panels in *a* and *b* correspond to the fluorescence traces marked with green circles, respectively. The total fluorescence was determined within the white outlines created via thresholding of the respective bright field images (and is given in fluorescent units (FU)).

will devise a rigorous way to quantify this delay. Here, we only apply a simple thresholding procedure to extract an apparent lag time. Using an intensity threshold of 2.5×10^4 fluorescence units, we determine an apparent lag time of 16 ± 2.5 min at 0.2% arabinose and a more substantial delay of 34 ± 10 min at 0.01% arabinose. In the latter case, ~10% of the bacteria do not show any fluorescence within our time window of 70 min.

With the sudden increase of the external arabinose concentration at $t = 0$ min, a cascade of biochemical processes is triggered, culminating in the fluorescent output signal measured in our experiment. To narrow down the origin of the stochasticity in the apparent lag time, we need to analyze the individual steps in this cascade. For this analysis, it is useful to separate the system into two distinct modules, an uptake module and a GFP expression module, as depicted in Fig. 3 *a*. The uptake module not only comprises arabinose import (represented here by an effective uptake protein, Upt, that subsumes transport by AraE and AraFGH) but also includes the positive feedback of arabinose on the uptake protein. The expression module turns on the production of the output signal, when internal arabinose reaches a threshold level (26). The delay time τ_D that is required to reach this threshold is solely determined by the uptake module. However, GFP fluorescence does not follow promoter activation instantaneously. Instead, the processes of transcription, translation, and GFP maturation depicted in Fig. 3 *b* also generate a dynamical delay and thereby contribute to the apparent delay estimated above. To quantitatively estimate the intrinsic delay τ_D and its statistics, we now scrutinize the expression module in detail.

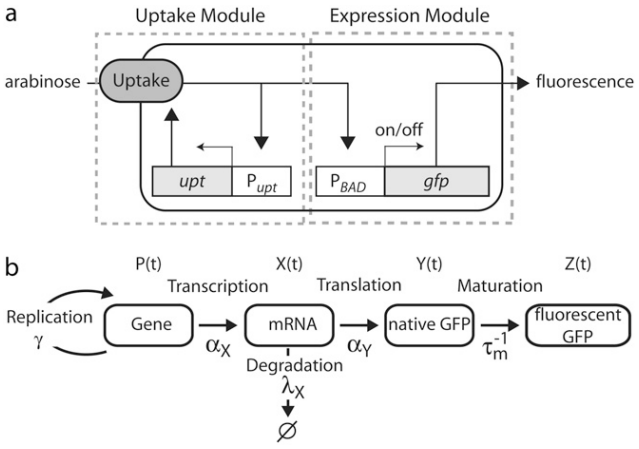


FIGURE 3 (*a*) The arabinose utilization system can be dissected into an arabinose uptake module (*left*) and a gene expression module (*right*). The gene expression module is turned on, if the internal arabinose level exceeds the threshold required for activation of the P_{BAD} promoter. The stochastic model for the uptake module comprises arabinose import by a heuristic uptake protein and the positive feedback of arabinose on the synthesis of the uptake protein (see Appendix B for all details). The model for the expression module encompasses the processes depicted in panel *b* and describes the accumulation of total fluorescent GFP per cell (see Appendix A for the deterministic rate equations).

Quantitative characterization of the expression module

GFP maturation time

A significant portion of the dynamic delay of the expression module is incurred by GFP maturation, the process whereby the folded protein becomes fluorescent. The rate-limiting reaction is an oxidation with a time constant of several minutes up to several hours (27), depending on the variant of the protein and possibly on the organism. However, for our present purpose, we not only need the average time constant, but also need to know whether there is a large cell-to-cell variation associated with the maturation process. With our microfluidic setup, we can directly probe this cell-to-cell variation experimentally, under the same conditions as in the induction experiments. First, we induce bacteria with 0.2% arabinose and then inhibit protein synthesis in situ by flushing the channel with the antibiotic chloramphenicol. The resulting fluorescence trajectories cease to increase ~15 min after the addition of the antibiotic (see Fig. 4 *a* for a few representative trajectories). Following the rationale established in Gordon et al. (25), this behavior reflects the maturation dynamics of the remaining, nonfluorescent GFPs. The distribution of time-constants τ_m of GFP maturation shown in Fig. 4 *b* was obtained from exponential fits to 77 single-cell time series (*solid lines* in Fig. 4 *a*). We find an average maturation time of $\tau_m = 6.5$ min and a standard deviation of 0.6 min, i.e., a cell-to-cell variation of only ~10%.

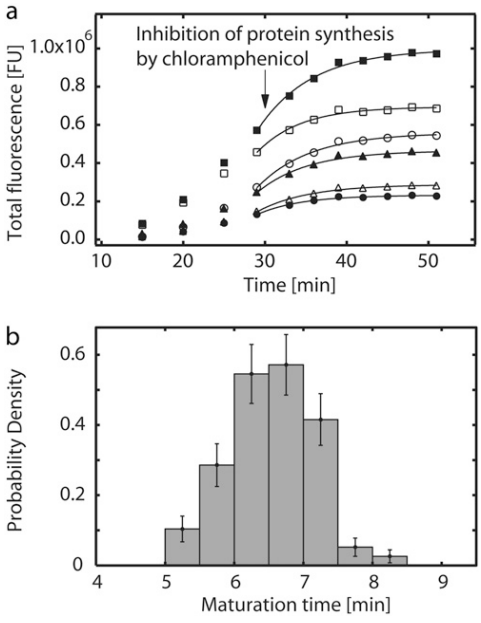


FIGURE 4 GFP maturation kinetics in single cells. In panel *a*, GFP expression was induced with 0.2% arabinose at $t = 0$ min and protein synthesis was inhibited by addition of 200 μg/ml chloramphenicol at $t = 30$ min, as indicated by the arrow. Exponential fits to the fluorescent time series (*solid lines*) yield the maturation-time distribution in panel *b*. The statistics was obtained from 77 cells.

Our finding of a relatively small cell-to-cell variation suggests that the maturation process is largely independent of the internal state of the cell in *E. coli*. This appears plausible, given that the oxidation reaction does not depend on intracellular components (27). For comparison, measurements of the maturation times of YFP and CFP in yeast (25) found considerably longer maturation times of ~40 min, but only a slightly larger relative cell-to-cell variation (15–20%). Moreover, from in vitro measurements of various YFP variants, oxidation timescales as low as 2–8 min were determined (28), indicating that the rapid maturation time detected in our experiment is conceivable in vivo.

Gene copy number

Since our GFP reporter is encoded on a plasmid, the average copy-number of the plasmid and its cell-to-cell variation are important properties of the expression module. The plasmid pBAD24 has an average copy number comparable to pUC (29), which is present in ~55 copies per cell (30). Assuming plasmid production and dilution with constant rates, we expect Poissonian fluctuations of $\sim\sqrt{55} \approx 7.5$ plasmids (13%). In similar plasmids, ColE1 and R1, negative feedback is known to reduce the copy-number variations below the Poisson limit (31). This may also apply to pBAD24, which would make the variation even less significant. We expect that the plasmid copy number grows proportional to the volume of the cell, such that the concentration of plasmids remains constant. Hence, we will assume that the rate γ of gene replication in Fig. 3 *b* equals the rate of volume expansion of the cells.

Cell growth

As the above discussion of the gene copy number shows, the distribution of growth rates is another characteristic affecting the quantitative properties of the expression module. We analyzed the growth of individual cells in the microfluidic channel by recording the time-evolution of their area detected under the microscope. Since the rod-shaped *E. coli* cells grow mainly along their principal axis (see image panels in Fig. 2), the growth rate of the cell area is a proxy for the growth rate by cell volume. From exponential fits (32) to 84 time series of the cell area we found a distribution of time constants for cell growth with an average of 50 min and a standard deviation of 6 min. Hence, the cell-to-cell variations of the growth rate are also relatively small. This result indicates that the microchemical conditions in our channel are sufficiently constant to guarantee a reproducible growth state of the cells. We also found that the doubling time was independent of the arabinose concentration, consistent with the fact that in this strain arabinose cannot be catabolized and used as an energy source.

mRNA half-life and protein expression rate

Finally, the dynamics of the expression module is dependent on the rate constants for *gfp* expression and mRNA degradation. Average mRNA half-lives were determined for most of *E. coli*'s genes (33) and are typically in the range 3–8 min. The work of Smolke et al. (34) indicates that the population-averaged half-life of *gfp* mRNA is in the same range; for our analysis below, we will assume an average half-life of 6 min. In contrast, there is currently no report on the cell-to-cell variation of *gfp* mRNA half-lives. We expect that such a variation would mainly be produced by cell-to-cell variations of RNase abundance and other components required for transcript turnover. These components have been shown to vary with the growth rate (35). Since the growth rate varies only by ~10% from cell to cell in our experiment (see above), we estimate the relative cell-to-cell variations of mRNA half-life to be similar. This may be an overestimate, since the degradation machinery negatively autoregulates its own expression (36), a mechanism known to reduce gene expression noise (37).

The protein expression rate has been quantified experimentally at the single-cell level for the P_R promoter of phage λ, and substantial cell-to-cell variations of ~35% were determined (38). These large relative differences likely stem from cell-to-cell variations in global cellular components such as RNA polymerases or ribosomes. We expect similar variations for GFP expression from the P_{BAD} promoter.

Distribution of GFP expression rate and intrinsic delay time

Given the above characterization of the expression module, we can now construct a simple quantitative model for its dynamic response, and then use this model to extract the intrinsic delay τ_D. The smooth shape of the time series in Fig. 2 suggests that the dynamics of individual cells follows a rather deterministic fate, while the differences between the cells stem from cell-to-cell variation of the reaction rates. Therefore, we use a deterministic rate equation model to describe the expression dynamics within a single cell, but allow for cell-to-cell variation in the model parameters. This model follows the reaction scheme depicted in Fig. 3 *b*: Transcription of *gfp* mRNA from the promoter P_{BAD} is turned on at $t = \tau_D$ and then remains constant at rate α_x. However, the number of plasmids (and hence gene copies) increases with rate γ, which equals the cell-doubling rate, so that the plasmid copy number P remains stable in the bacterial population. We denote the mRNA degradation rate by λ_x, and the translation and maturation rates of GFP by α_y and τ_m^{-1}, respectively (see Appendix A for details).

Within this model, the time-evolution of the total number of fluorescent GFP molecules in a cell, $Z(\tau)$, is described by the expression

$$Z(\tau) = \alpha_p \left(\frac{(\gamma + \lambda_x) e^{-\tau/\tau_m}}{(\gamma + \tau_m^{-1})(\lambda_x - \tau_m^{-1})} + \frac{\tau_m^{-1} e^{-\lambda_x \tau}}{\lambda_x (\tau_m^{-1} - \lambda_x)} + \frac{\tau_m^{-1} e^{\gamma \tau}}{\gamma (\gamma + \tau_m^{-1})} \right) - Z_0, \quad (1)$$

where $\tau = t - \tau_D$ is the time after transcription is switched on, $\alpha_p \equiv P\alpha_x\alpha_y/(\gamma + \lambda_x)$ is a lumped constant giving the protein synthesis rate in fluorescence units per minute [FU/min], and Z_0 is a constant determined by the initial conditions. Here, the first two terms in parentheses describe transients associated with the equilibration of the GFP maturation process and the mRNA degradation reaction, respectively, i.e., their contributions decay exponentially with time constants τ_m and λ_x^{-1}. In the long-time limit, the last, exponentially increasing term is dominant. It reflects the constant protein production from an exponentially growing number of plasmids, and describes the long-time behavior of the total fluorescence per cell. However, since we study the dynamics of gene expression during the first cell cycle after induction, all terms, including the transients, are relevant.

From the previous section, we conclude that the parameter α_p, comprising the plasmid copy number and the protein expression rate, captures most of the cell-to-cell variation within the expression module. To fit the model in Eq. 1 to the single-cell induction kinetics, we therefore fixed the remaining parameters to their population-averaged values. Hence, in the optimization procedure of the fit, we only allow the adjustment of α_p and the uptake-induced delay τ_D, which we sought to extract. Note that this choice fixes all relevant timescales governing the dynamics in Eq. 1 and the free parameters only impose shifts in the onset (τ_D) and in the absolute magnitude (α_p) of *gfp*-expression.

We fitted the time series of cells induced with various levels of arabinose (0.2%, 0.05%, 0.02%, and 0.01%). A few representative fitted curves for the highest and lowest concentration are plotted in Fig. 2 as solid lines. The resulting histograms for the delay time are shown in Fig. 5 *a*. For the lowest arabinose level (0.01%, *upper panel*) we find that the delay times are distributed between 5 and 50 min with a mean and standard deviation of $\langle\tau_D\rangle = 23$ min and $\sigma_{\tau_D} = 10$ min, respectively. With increasing arabinose concentration, both the mean and the standard deviation of the delay time distribution decrease gradually, until at the highest arabinose level (0.2%, *lower panel*), a distribution with $\langle\tau_D\rangle = 4.1$ min and $\sigma_{\tau_D} = 2.2$ min is reached.

To test whether there is a relationship between the delay time and the protein synthesis rate, we calculated their cross-correlation coefficients for all inducing arabinose levels (see Fig. 6 *a*). Only in the case of 0.02% arabinose a slight anti-correlation was detected, whereas for all other concentrations the correlation coefficient is close to zero (*p*-values for finding the observed correlation coefficients by chance in an uncorrelated sample: 0.68 for 0.01% ara, 0.03 for 0.02% ara, 0.73 for 0.05% ara, and 0.72 for 0.2% ara). We also find that the distribution of *gfp*-expression rates itself does not vary systematically with the inducing arabinose concentration, and all distributions fall on top of each other when rescaled by their mean values, see Fig. 6 *b* (pairwise Kolmogorov-Smirnov tests yield significance levels between 0.57 and 0.97 for the null hypothesis that the data sets are drawn from the same underlying distribution). In summary, the low correlations between

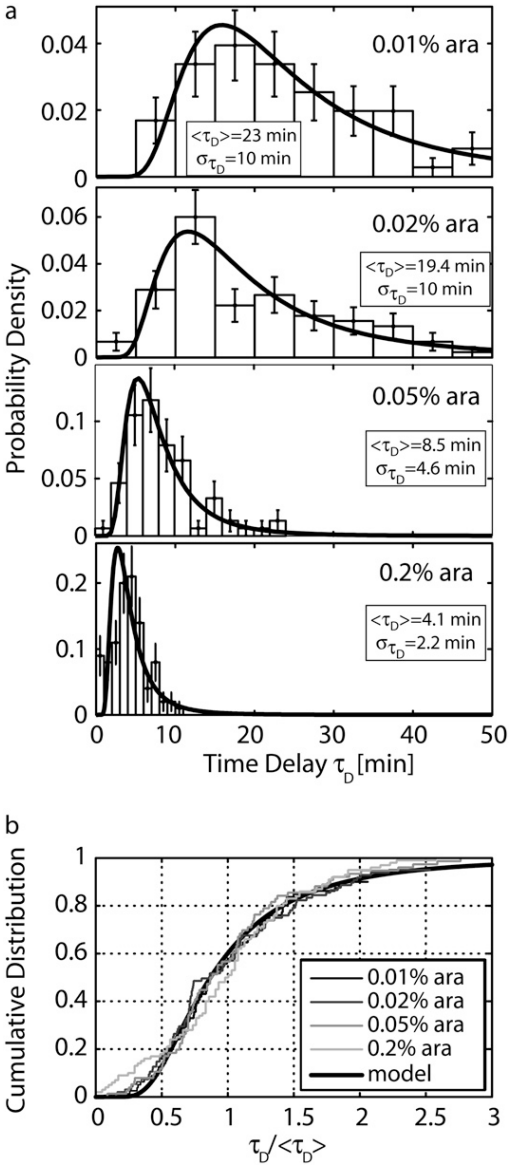


FIGURE 5 (*a*) Histograms of the time delay τ_D for varying external arabinose concentrations, as determined from the fits of Eq. 1 to the fluorescence time series. The mean $\langle\tau_D\rangle$ as well as the standard deviation σ_{τ_D} gradually decrease for increasing arabinose levels. Note that for 0.01% arabinose, ~10% and for 0.02% arabinose, ~5% of the cells did not turn on gene expression within our experimental time window. Therefore the extracted means and standard deviations (see *insets*) constitute slight underestimates in these cases. The solid lines are fits of the analytical delay time distributions (Eq. 3) to the data (for details, see text). The statistics was obtained from 71 cells at 0.01%, 90 cells at 0.02%, 76 cells at 0.05%, and 101 cells at 0.2% arabinose. (*b*) Cumulative distributions of the delay times rescaled to their mean values $\langle\tau_D\rangle$. A two-sample Kolmogorov-Smirnov test indicates that all rescaled distributions are likely to be drawn from the same underlying probability distribution. The *p*-values of the individual pairs are 0.25 for 0.01% and 0.02% arabinose; 0.87 for 0.01% and 0.05% arabinose; 0.54 for 0.01% and 0.2% arabinose; 0.90 for 0.02% and 0.05% arabinose; 0.08 for 0.02% and 0.2% arabinose; and 0.67 for 0.05% and 0.2% arabinose. The analytical prediction is shown for $\mu = 3.8$, $b = 30$, and $\tau_0 = 2100$ min (*bold line*).

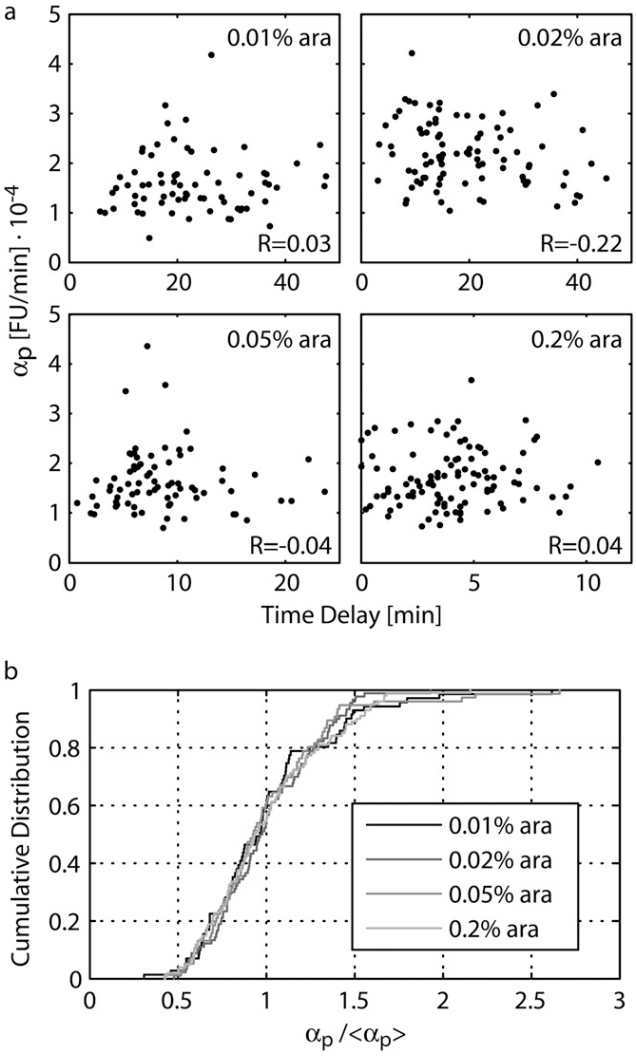


FIGURE 6 (*a*) Correlations between delay time and protein synthesis rate α_P. The scatter plots display small correlation coefficients R, and the respective p-values for observing these correlations by random chance are 0.68 for 0.01%, 0.035 for 0.02%, 0.73 for 0.05%, and 0.72 for 0.2% arabinose. (*b*) The cumulative distributions of the protein synthesis rate α_P were rescaled to their mean values $\langle\alpha_P\rangle$ to exclude sample-to-sample variations of the mean. Importantly, we found no correlations between $\langle\alpha_P\rangle$ and the inducing arabinose concentration. A two-sample KS-test shows that all rescaled distributions of α_P are compatible with each other. The significance levels of the pairwise KS-tests varied between 0.57 and 0.97.

τ_D and α_P on the one hand, and the independence of α_P on the inducing arabinose level on the other hand, suggest that the uptake and the expression module are indeed functionally separate. Note that our experimental approach with time-lapse fluorescence microscopy was crucial for these results, which would have been impossible to obtain with flow cytometry.

Stochastic model for the uptake module

Next, we want to assess whether the extracted delay time distributions of Fig. 5 *b* may be causally linked to a broad variation in the number of uptake proteins. We approach this question with the help of a simple stochastic model for the uptake-module depicted in Fig. 3 *a*. The model is useful in three respects:

1. It serves us to illustrate the mechanism whereby stochastic expression of the uptake protein genes can produce a broad distribution of delay times. We will see that according to this mechanism, the delay time distributions for different inducer concentrations should be related by simple linear rescaling of the time axis. Thus, we will test for this signature of the mechanism in our experimental data.
2. Since most model parameters are strongly constrained by literature values, we can test whether an interpretation of our data based on the stochastic model is consistent with these constraints.
3. Independent of the precise choice of parameter values, which affect the average delay time and its standard deviation, the model predicts a certain shape for the delay time distribution. We will test whether this shape is compatible with our data.

There are two distinct transport systems for arabinose uptake, AraE and AraFGH. However, the two systems are coupled, and it was found that arabinose uptake can effectively be described as a single Michaelis-Menten process (39). In the sketch of Fig. 3 *a*, this combined transport system is represented by a single gene *upt*. In addition to the transport, the uptake module of Fig. 3 *a* comprises the activation of AraC by internal arabinose, the subsequent stimulation of transcription by the activated complex, and the translation into functional uptake protein. Within our stochastic model for the uptake module, we describe and simulate all of these processes in standard ways (see Appendix B for details).

Fig. 7, *b* and *c*, show the simulated time-evolution of the level of uptake proteins and the level of internal arabinose upon induction with 0.01% external arabinose for a few representative simulation runs. These trajectories illustrate the mechanism leading to a broad distribution of delay times within our model: Internal arabinose initially accumulates approximately linearly in time, and the accumulation accelerates only after reaching the effective arabinose threshold of $a_0 \approx 50\ \mu$M for activation of the *araBAD* and *upt* promoters, which is indicated by the solid horizontal line in Fig. 7 *c*. The time delay, τ_D, caused by the uptake module is the time required for the internal arabinose concentration to reach this threshold level. The rate of arabinose import, given by the slope in Fig. 7 *c*, is proportional to the number of uptake proteins n in Fig. 7 *b*. If arabinose import is fast compared to the timescale of changes in the protein abundance, the delay time is given by the simple relation $\tau_D = a_0/(v_0 n)$, where the arabinose uptake rate per uptake protein, v_0, depends on the external arabinose concentration. Thus, the distribution of uptake proteins in Fig. 7 *a* directly determines the distribution of import rates, which in turn are inversely proportional to the

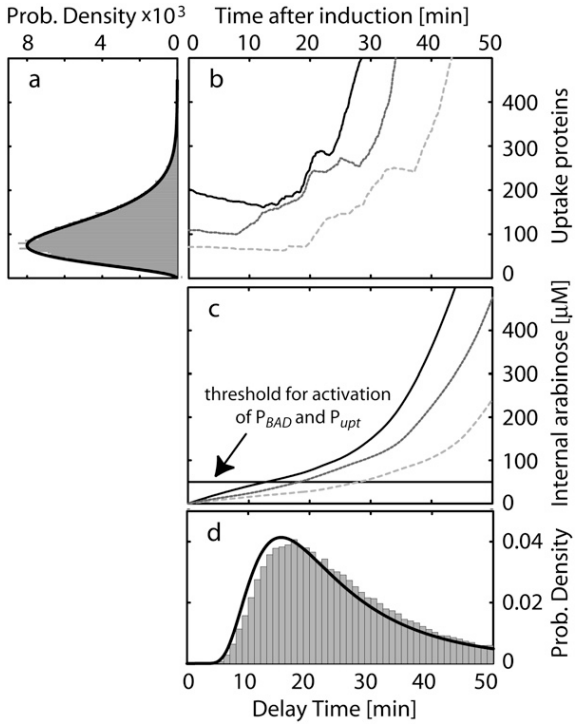


FIGURE 7 Illustration of the stochastic arabinose uptake mechanism at 0.01% external arabinose (simulation). The three representative time-courses of arabinose uptake proteins in panel *b* and internal arabinose in panel *c* illustrate that the rate of arabinose uptake is proportional to the amount of uptake protein present at a given time. Once the internal threshold for activation of the promoters P_{BAD} and P_{upt} is reached, the positive feedback gets activated and is visible as the kinks in panels *b* and *c*. The delay time distribution in panel *d* (*shaded bars*) is obtained by measuring the time to reach this threshold. If the uptake proteins decay much slower than the typical time required to reach the threshold (adiabatic limit), the delay time distribution in panel *d* can be related to the steady-state distribution of uptake proteins at zero arabinose in panel *a*. The analytical predictions in panels *a* and *d* (*solid lines*) are shown for $\mu = 3.8$, $b = 30$, and $\tau_0 = 2100$ min (for details see text).

delay times, resulting in the distribution of delay times shown in Fig. 7 *d*.

A simple prediction of this mechanism is that an increase of the uptake velocity v_0 will reduce all delay times within a distribution of cells by the same factor. In other words, the delay time distributions for different arabinose levels (and hence different v_0) should fall on top of each other upon simple linear rescaling of the time axis (and restoring normalization). In Fig. 5 *b*, we test this prediction on our experimental time delay distributions. We find that after rescaling to the same mean value, the cumulative distributions are congruent with each other. This agreement is also quantitatively supported by pairwise Kolmogorov-Smirnov tests, which test whether the samples are likely to be drawn from the same underlying distribution (the legend to Fig. 5 shows the respective significance levels). Note that the linear scaling of the time axis with $1/v_0$ does not imply linear scaling with the arabinose level, since v_0 depends nonlinearly on the external arabinose level (see also further below).

To relate the experimentally observed shape of the distribution to the prediction of the stochastic model, we will now derive an analytical expression for the delay time distribution. To this end, we first consider only intrinsic noise and study the effect of extrinsic noise below. Before the addition of the inducer arabinose, expression of the uptake proteins is a completely random, unregulated process. Following the work of Berg (40) and under the assumptions stated in Appendix B, we find a steady-state distribution $P(n)$ for the number of uptake proteins n of the form

$$P(n) \approx \left(\frac{1}{1+b}\right)^{\mu}\left(\frac{b}{1+b}\right)^{n}\binom{\mu+n-1}{n}, \qquad (2)$$

which is sometimes referred to as a negative binomial. Here, the ratio $b = \nu_{\mathrm{p}}/\lambda_{\mathrm{m}}$ of the translation rate and the mRNA degradation rate corresponds to the typical number of proteins produced from a single mRNA and is also known as the burst size (41). The ratio $\mu = \nu_{\mathrm{m}}^{0}/\lambda_{\mathrm{p}}$ of the basal transcription rate and the protein dilution rate can be interpreted as a dimensionless burst frequency (the number of bursts within the lifetime of a protein). Both parameters determine the mean $\langle n\rangle = \mu b$ and the variance $\delta n^2 = \langle n\rangle\,(1 + b)$ of $P(n)$. Fig. 7 *a* shows the steady-state distribution $P(n)$ obtained from our stochastic simulations of the uptake module (*shaded histogram*), together with the analytical expression in Eq. 2 for the same rate constants. The excellent agreement suggests that the assumptions leading to Eq. 2 are all satisfied in the relevant parameter regime.

Next, we study the effect of extrinsic noise which leads to a variation of reaction parameters from cell to cell. An experimental characterization of extrinsic noise in *E. coli* (38) found a typical parameter variation of ~20%. When we adopt this level of extrinsic noise for all parameters in our stochastic simulations, the resulting protein distribution has a significantly larger standard deviation than the distribution in the absence of extrinsic noise, while the mean remains almost unchanged (see Supplementary Material, Fig. S1). However, the protein distribution in the presence of extrinsic noise is still well fitted by Eq. 2, with an increased effective burst size and a reduced effective burst frequency. Keeping this in mind, the following results can be generalized to the realistic scenario where extrinsic fluctuations are present.

To obtain an approximation for the delay time distribution, we assume that arabinose uptake is rapid compared to the typical timescale of changes in the protein abundance. In this adiabatic limit, the delay time is inversely proportional to the current protein abundance in each cell, i.e., $\tau_{\mathrm{D}} = \tau_0/n$, where $\tau_0 \equiv a_0/v_0$ is the time for a single uptake protein to accumulate arabinose to the threshold level a_0. With this relation, the steady-state uptake protein distribution (Eq. 2) leads to a delay time distribution of the form

$$Q(\tau_{\mathrm{D}}) \approx \frac{\tau_0}{\tau_{\mathrm{D}}^2}\left(\frac{1}{1+b}\right)^{\mu}\left(\frac{b}{1+b}\right)^{\tau_0/\tau_{\mathrm{D}}}\frac{\Gamma(\tau_0/\tau_{\mathrm{D}}+\mu)}{\Gamma(\tau_0/\tau_{\mathrm{D}}+1)\Gamma(\mu)}, \qquad (3)$$

where $\Gamma(x)$ is the gamma function. In Fig. 7 *d*, we compare this analytical prediction (*solid line*) to the stochastic simulation (*shaded bars*). The small deviation stems from the fact that the number of uptake proteins is not constant over the period of the time delay. Note that, indeed, if the protein dynamics is much faster than the characteristic time of arabinose uptake ($\lambda_p^{-1} \ll \tau_D$), every cell experiences simply the average abundance of uptake protein $\langle n \rangle$ and the delay time distribution approaches a sharply peaked function at $\sim\tau_D = \tau_0\langle n\rangle^{-1}$ (data not shown). In our case, $\lambda_p^{-1} \approx 70$ min is much larger than the average delay times, so that the assumption of a constant n is sufficiently accurate. The mean and variance of the delay time distribution can be approximated by

$$\langle\tau_D\rangle \approx \frac{\tau_0}{\langle n\rangle}\left(1+\frac{\delta n^2}{\langle n\rangle^2}\right) \approx \frac{\tau_0}{\mu b}\left(1+\frac{1}{\mu}\right),$$
$$\delta\tau_D^2 \approx \frac{\tau_0^2}{\langle n\rangle^2}\frac{\delta n^2}{\langle n\rangle^2} \approx \left(\frac{\tau_0}{\mu b}\right)^2\frac{1}{\mu}, \quad (4)$$

(see Appendix B). From these expressions, it is clear that the model has two key parameters, which together determine the mean and width of the delay time distribution: the time required to reach the internal arabinose threshold by a single protein burst, τ_0/b, and the burst frequency μ.

Now we test whether the shape of the delay time distribution predicted by the model is quantitatively consistent with our experimental distributions. To this end, we fit the model in Eq. 3 to the data in Fig. 5 *a* by varying the two key parameters identified above. The resulting fits (*solid lines*) display good agreement with the experimental data, as indicated by one-sample KS-tests under the null hypothesis that the samples are drawn from the analytical distribution. The significance levels are 0.50, 0.47, 0.77, and 0.07 for 0.01%, 0.02%, 0.05%, and 0.2% arabinose, respectively. Only in the case of 0.2% arabinose does the test point to a significant difference between the theoretical and experimental distribution. However, for this concentration the estimated delay times are very short, such that the error of the estimation itself is likely to account for the deviations. Note that the two-parameter fit guarantees that the mean and standard deviation of the experimental and theoretical distribution will match. However, the fact that the shape of the distributions shows excellent agreement is a nontrivial result, suggesting that the discussed delay mechanism can indeed explain our observations.

Finally, we address the consistency of the parameter values. Fig. 8 shows the estimated parameters as a function of the external arabinose concentration. The timescale τ_0/b of arabinose accumulation in Fig. 8 *a* decreases monotonically as a function of external arabinose and saturates for large sugar abundances, whereas the burst frequency μ in Fig. 8 *b* is constant for all arabinose levels. This observation is consistent with the idea that the underlying protein distribution, characterized by μ and b, is independent of the externally provided sugar concentration, and that the differences in timing can be explained by shifts in the effective arabinose uptake velocity per uptake protein, v_0: By assuming simple Michaelis-Menten saturation kinetics for v_0, one expects that τ_0 scales inversely with the external arabinose concentration $[a_{ex}]$, i.e., $\tau_0 = a_0/v_{max}(1+K_m/[a_{ex}])$, where v_{max} denotes the maximal uptake velocity per uptake protein and K_m the Michaelis constant. This behavior is indeed found in Fig. 8 *a* (*inset*) and with the resulting values for v_{max}, K_m and a typical value of $b = 30$ for the burst factor (41), all parameters are compatible with the experimentally constrained ranges discussed in Appendix B.

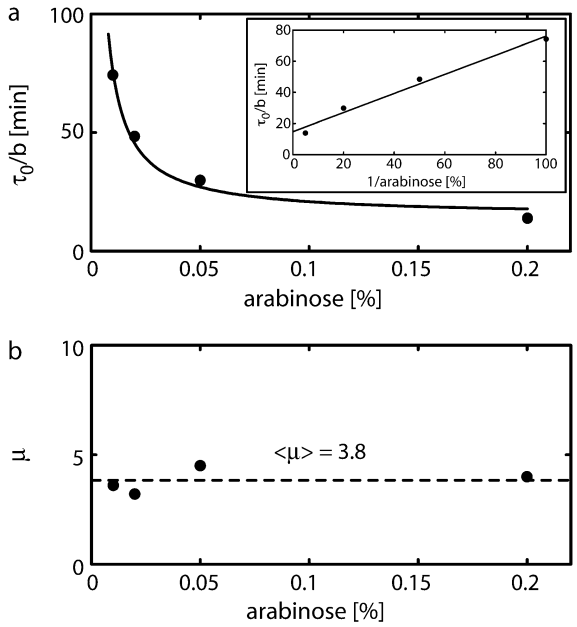


FIGURE 8 Estimated parameters as a function of external arabinose, as obtained from fits of the delay time distributions in Fig. 5 *a*. The timescale of arabinose accumulation τ_0/b in panel *a* decreases monotonically with the inducing arabinose concentration, as expected from saturating Michaelis-Menten kinetics of the uptake proteins; compare the Lineweaver-Burk plot (*inset*) for the scaling with the inverse arabinose concentration. In contrast, the burst frequency μ in panel *b* is constant for all arabinose levels. This is consistent with our central assumption that the underlying uptake protein distribution responsible for the heterogeneous timing is independent of the inducing arabinose concentration.

DISCUSSION

We studied the expression dynamics during induction of the bistable arabinose utilization system in single *E. coli* cells using quantitative time-lapse fluorescence microscopy. Upon addition of arabinose, we observed a characteristic time delay before the cells switched from a state of basal expression to a state of high expression of the *ara* regulon. This typical duration of the delay exhibited a systematic dependence on the externally supplied arabinose concentration: At a saturating arabinose level, we found rapid induction within all cells of the culture, whereas with decreasing levels, we detected a significant broadening and shift of the delay time distribution

function. To characterize the cell-to-cell variability in the cellular response, we dissected the system into an uptake module with stochastic behavior, and an expression module which displays virtually deterministic behavior in individual cells. We first studied the expression module, in particular by measuring the cell-to-cell distribution of the GFP maturation time. To the best of our knowledge, this constitutes the first measurement of a maturation time distribution in bacteria. We then developed a hybrid deterministic/stochastic theoretical model to analyze our experimental data. The model is based on the assumption that the initial basal expression of the arabinose transporters determines the rate of arabinose uptake. Adopting the approach of Berg, we find an analytic expression for the distribution of transporter proteins and the distribution of delay times. The theory consistently fits the shape of the experimental delay time distributions for various inducer concentrations. Hence our data support a previous conjecture by Siegele and Hu (12), according to which the delay time distribution is causally linked to the distribution of uptake proteins in the absence of the inducer. To corroborate our model even further, it would be interesting to control the level of transporter proteins independently, e.g., by using an inducible promoter that is independent of arabinose. Also, it remains an open question how the two transport systems are coupled. It appears that the high-affinity low-capacity transporter *araFGH* and the low-affinity high-capacity transporter *araE* are orchestrated to respond like a single protein. A similar analysis to ours using knockout mutants in one of the two transport systems could shed light on this matter.

In general, we determined the dynamic response of bacteria to an external change of food conditions. Since such decisions are of vital importance to living systems, we can speculate about their impact on the fitness of a bacterial population. The observed heterogeneous timing in gene induction may simply be a fortuitous consequence of the evolutionary process that shaped the arabinose utilization system in *E. coli*. Alternatively, it may be beneficial for a bacterial colony, if the individual cells respond at different times when arabinose suddenly becomes available in modest amounts. Note that in our experiments with the *araBAD* deficient strain, even the lowest arabinose level, if maintained over a long time, ultimately induces the *ara* system in almost all cells. However, for a wild-type strain in an environment where arabinose availability may fluctuate, temporal disorder of gene induction could provide selective advantages for the colony as a whole. For instance, it might be beneficial to prevent costly synthesis of the arabinose system in all cells when the sugar level is only moderate and may soon be depleted. Our analysis indicates that the delay time distribution of the system can be readily tuned over evolutionary timescales, by adjusting the burst frequency and burst size of the uptake proteins. In the future, it will be interesting to further explore the possible connections between the system design in individual cells and the biological function at the population level.

APPENDIX A: DETERMINISTIC GFP EXPRESSION MODEL

To extract the intrinsic time delay τ_D from our single cell expression data, we employ a simple deterministic model that follows the scheme depicted in Fig. 3 *b*. We assume that the transcription rate from the promoter P_{BAD} is zero until the internal arabinose threshold for activation of P_{BAD} is reached at $t = \tau_D$. Then, the promoter activity jumps to its maximal value α_x. The corresponding rate-equations for the total abundance of plasmids (P), *gfp* mRNA (X), immature GFP protein (Y), and mature GFP protein (Z) per cell, are

$$\partial_t P = \gamma P$$
$$\partial_t X = \alpha_x P - \lambda_x X$$
$$\partial_t Y = \alpha_y X - \tau_m^{-1} Y$$
$$\partial_t Z = \tau_m^{-1} Y.$$

with the cell-doubling rate γ and the rate for transcription α_x, translation α_y, maturation τ_m^{-1}, and mRNA degradation λ_x. Note that the model does not include dilution due to cell growth, since we measured the total fluorescence per cell in our experiments. Therefore the number of plasmids (number of gene copies) increases exponentially in time, keeping the number of genes per volume constant. Solving these equations for $Z(\tau)$ leads to Eq. 1 in the main text.

APPENDIX B: STOCHASTIC MODEL FOR ARABINOSE UPTAKE

The arabinose uptake module, see Fig. 3 *a*, includes the processes for the uptake of arabinose as well as transcription, translation, and turnover of uptake proteins. In the following we describe the chemical reactions included in the stochastic simulations used to generate Fig. 7 and Fig. S1. We then derive an analytical approximation for the delay time distribution and discuss the experimental constraints on the model parameters.

Arabinose uptake

Comparison of arabinose uptake in wild-type strains with *araE* and *araFGH* deletion strains revealed that the two transporters do not operate independently (39). Instead, arabinose transport was best described by a single Michaelis-Menten function. Our model reflects this behavior of the wild-type strain through the use of a single effective uptake protein (referred to as Upt) for arabinose import,

$$a_{ex} + \mathrm{Upt} \overset{K_m}{\leftrightarrow} a_{ex} \cdot \mathrm{Upt}$$
$$a_{ex} \cdot \mathrm{Upt} \overset{v_{max}}{\rightarrow} a + \mathrm{Upt}.$$

The uptake protein binds external arabinose a_{ex} with dissociation constant K_m and, once bound, translocates it to the cytoplasm at rate v_{max}. The effective uptake velocity per uptake protein is hence $v_0 = v_{max}[a_{ex}]/(K_m + [a_{ex}])$. Cytoplasmic arabinose is denoted by a.

Transcriptional regulation

The P_{BAD} promoter in the *ara*-regulon is one of the best characterized bacterial promoters: In the presence of internal arabinose, AraC stimulates transcription from P_{BAD}, while AraC represses transcription by formation of a DNA loop in the absence of arabinose (11). When exceeding an arabinose threshold of $a_0 \approx 50\ \mu$M, the promoter activity of P_{BAD} increases cubically with the internal arabinose concentration (26). In contrast to the detailed studies on P_{BAD}, less is known about the promoter activity function of the promoters P_E and P_{FGH}, which regulate expression of the transport proteins.

Both promoters are also induced by internal arabinose, but lack an upstream AraC-binding site (required for DNA looping) and are not repressed in the absence of arabinose. Consequently, their basal expression level is higher than for P_{BAD} and the fold-change is reduced from ~400 for P_{BAD} to ~150 for P_E and P_{FGH} (42). However, the detailed promoter activity as a function of internal arabinose is not known for these promoters. Apart from the lack of the AraC binding site required for DNA looping, the promoters P_E and P_{FGH} display a high similarity to P_{BAD} (43). Therefore we model transcriptional regulation of the uptake proteins by introducing a heuristic promoter P_{upt}, which has the same characteristics as P_{BAD}, but lacks the repression in the absence of arabinose. To reproduce the cubic increase of the promoter activity function of P_{BAD} we allow three arabinose molecules to bind AraC with dissociation constant K_C. This activated complex binds the promoter P_{upt} with dissociation constant K_P and thereby switches the transcription rate from its basal rate ν_m^0 to its maximal rate ν_m. The chemical reactions for transcriptional regulation are

$$
\begin{aligned}
&3\,a + C \overset{K_C}{\leftrightarrow} a_3 \cdot C\\
&a_3 \cdot C + P_{upt} \overset{K_P}{\leftrightarrow} a_3 \cdot C \cdot P_{upt}\\
&P_{upt} \overset{\nu_m^0}{\rightarrow} P_{upt} + m\\
&a_3 \cdot C \cdot P_{upt} \overset{\nu_m}{\rightarrow} a_3 \cdot C \cdot P_{upt} + m.
\end{aligned}
$$

Here the concentration of AraC molecules [C] is a variable that changes little over time (43) and is therefore assumed to be a constant parameter in our model. In steady state, the probability for finding the promoter P_{upt} in a transcriptionally activated state is a Hill function of the internal arabinose concentration, $[a_3 \cdot C \cdot P_{upt}] = [a]^3/(K_C K_P/[C] + [a]^3)$. We define the effective arabinose threshold for activation of P_{upt} as $K_{upt} \equiv (K_C K_P/[C])^{1/3}$.

Translation and turnover

mRNA is translated into functional uptake protein at rate ν_p and gets degraded at rate λ_m. In contrast, the uptake proteins and arabinose are only diluted by cell growth at doubling rate γ:

$$
\begin{aligned}
&m \overset{\nu_p}{\rightarrow} m + \mathrm{Upt}\\
&m \overset{\lambda_m}{\rightarrow} \emptyset\\
&\mathrm{Upt} \overset{\gamma}{\rightarrow} \emptyset\\
&a \overset{\gamma}{\rightarrow} \emptyset.
\end{aligned}
$$

Delay time distribution

Following Berg (40), we derive an analytical approximation for the delay time distribution of our stochastic model. In the absence of arabinose, transcription of the gene for the uptake protein takes place at its basal rate ν_m^0. Neglecting operator state fluctuations (44), the probability to observe m transcription events up to time t follows a Poisson distribution

$$
\mathrm{P}(m|\nu_m^0 t) = \frac{(\nu_m^0 t)^m}{m!} e^{-\nu_m^0 t},
$$

with mean and variance $\nu_m^0 t$. In the limit of short mRNA lifetime λ_m^{-1} compared to the protein lifetime λ_p^{-1}, one can assume instantaneous, geometrically distributed protein bursts from each mRNA molecule. This implies that the probability that m mRNA molecules produce n proteins follows a negative binomial distribution

$$
\mathrm{NB}(n|m,b) = \left(\frac{1}{1+b}\right)^m \left(\frac{b}{1+b}\right)^n \binom{\mu+n-1}{n},
$$

where the burst size $b \equiv \nu_p/\lambda_m$ is the average number of proteins produced from one mRNA molecule. Hence, the probability to produce n proteins up to time t is the weighed sum of negative binomials $P(n|\nu_m^0 t, b) \equiv \sum_m \mathrm{P}(m|\nu_m^0 t) \times \mathrm{NB}(n|m,b)$. Setting t equal to the protein lifetime λ_p^{-1} yields the steady-state distribution of proteins, and for large $\mu \equiv \nu_m^0/\lambda_p$ we can replace the Poisson distribution by a δ-function located at $m = \mu$, leading to Eq. 2 in the main text. Applying the transformation rule $Q(\tau_D) = |\mathrm{d}n(\tau_D)/\mathrm{d}\tau_D| P(n)$ yields the delay time distribution in Eq. 3 and the moments $\langle\tau_D\rangle$ and $\delta\tau_D^2 = \langle\tau_D^2\rangle - \langle\tau_D\rangle^2$ are determined by the integrals

$$
\begin{aligned}
\langle\tau_D\rangle &= \int d\tau_D \tau_D Q(\tau_D) = \int dn \frac{\tau_0}{n} P(n), \quad \text{and}\\
\langle\tau_D^2\rangle &= \int d\tau_D \tau_D^2 Q(\tau_D) = \int dn \frac{\tau_0^2}{n^2} P(n).
\end{aligned}
$$

Here expansion of the integrands up to second order in $\delta n = n - \langle n\rangle$ brings us to the expressions in Eq. 4.

Parameter values

The effective arabinose threshold $K_{upt} \approx 50\ \mu$M and the promoter binding constant $K_P = 10$ nM are chosen similar to the parameters of P_{BAD} (26,45). This choice determines the ratio $K_C/[C] = K_{upt}^3/K_P$ (see above) and by choosing a typical value of [C] = 100 nM we obtain $K_C = 10^6\ \mu\mathrm{M}^3$. For the maximal promoter activity we set a typical value for the promoters in the *ara*-regulon, $\nu_m = 5$ mRNA/min, which was derived from the mRNA steady-state levels reported in Johnson and Schleif (43). With a promoter fold-change of 150 similar to P_E and P_{FGH} (42), the basal transcription rate is expected to be $\sim\nu_m^0 \approx 0.03$ mRNA/min . From our fits of Eq. 3 to the experimental delay time distributions we obtained an average value of $\mu = \nu_m^0/\lambda_p = 3.8$. With our protein dilution rate of $\lambda_p = \gamma = \ln(2)/(50\ \mathrm{min})$ (from our measurement of the growth rate, see main text), this yields a basal expression rate of $\nu_m^0 \approx 0.05$ mRNA/min—in good agreement with the biochemical constraints stated before. The mRNA degradation rate λ_m is set according to a half-life of 2 min (43), allowing us to adjust the translation rate ν_p to match a typical burst factor of $b = 30$ (41). The K_m for arabinose uptake is in wild-type cells at ~50 μM (39), and the maximal uptake rate per uptake protein, v_{max}, can be estimated from bulk measurements in which the uptake rate per total cellular dry mass was determined (39). By assuming a dry mass of 3×10^{-13} g per cell (46) and $\sim 10^3$–10^4 uptake proteins per cell (47), we end up with $v_{max} = 200$–2000 arabinose molecules/protein/min. From a Lineweaver-Burk fit to the data in Fig. 8 *b* we obtained $v_{max} \approx 120$ molecules/protein/min and $K_m = 2.8$ mM. While the value for v_{max} is compatible with the biochemical constraints, our K_m differs by two orders of magnitude from the previously reported value of 50 μM (39). For such a small Michaelis constant, all arabinose concentrations used in our experiments would saturate the uptake system completely and hence there should be no difference in timing of gene induction. However, the experimental conditions of Daruwalla et al. (39) differ from ours; in particular, the proton gradient between periplasm and cytoplasm, which drives the arabinose/H^+ symport by AraE, is limited by oxygen availability (48). For the case of the lactose/H^+ symporter LacY, it has been shown that a reduced proton gradient leads to an increase of the apparent K_m (49). Hence, oxygen limitation in our microfluidic setup could explain the observed discrepancy.

Stochastic simulations

Although in the rate equations above only the equilibrium constants are depicted, we took for the dynamical simulations all association and dissociation processes explicitly into account. As a conservative assumption, all association rates were chosen 10-fold smaller than the diffusion-limited on-rate of 2 $\mathrm{nM}^{-1}\ \mathrm{min}^{-1}$ for a typical transcription factor in *E. coli* (50) and the dissociation rates were adjusted according to the respective equilibrium constant. The trajectories in Fig. 7, *b* and *c*, correspond to single kinetic

Monte Carlo simulations (51) for 0.01% external arabinose. The protein and delay-time distributions in Fig. 7, *a* and *d* (*solid lines*), were obtained from 5×10^4 independent simulation runs with the same parameters.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

We are grateful to R. Heermann for construction of the plasmid. We thank T. Hwa for helpful discussions and M. Leisner for careful reading of the manuscript.

This work was supported by the LMUinnovativ project ''Analysis and Modeling of Complex Systems''. J.A.M. acknowledges funding by the *Elitenetzwerk Bayern*. Author contributions: J.A.M. carried out the experiments. G.F. performed the simulations and analytical calculations. All authors designed the research and wrote the article.

REFERENCES

1. Kaern, M., T. C. Elston, W. J. Blake, and J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6:451–464.
2. Süel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*. 440:545–550.
3. Leisner, M., K. Stingl, J. O. Rädler, and B. Maier. 2007. Basal expression rate of comK sets a switching-window into the K-state of *Bacillus subtilis*. *Mol. Microbiol.* 63:1806–1816.
4. Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler. 2004. Bacterial persistence as a phenotypic switch. *Science*. 305:1622–1625.
5. Paliwal, S., P. A. Iglesias, K. Campbell, Z. Hilioti, A. Groisman, and A. Levchenko. 2007. MAPK-mediated bimodal gene expression and adaptive gradient sensing in yeast. *Nature*. 446:46–51.
6. Isaacs, F. J., J. Hasty, C. R. Cantor, and J. Collins. 2003. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA*. 100:7714–7719.
7. Thattai, M., and A. van Oudenaarden. 2004. Stochastic gene expression in fluctuating environments. *Genetics*. 167:523–530.
8. Kussell, E., and S. Leibler. 2005. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*. 309: 2075–2078.
9. Novick, A., and M. Weiner. 1957. Enzyme induction is an all-or-none phenomenon. *Proc. Natl. Acad. Sci. USA*. 43:553–566.
10. Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden. 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*. 427:737–740.
11. Schleif, R. 2000. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet.* 16:559–565.
12. Siegele, D. A., and J. C. Hu. 1997. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. USA*. 94: 8168–8172.
13. Carrier, T. A., and J. D. Keasling. 1999. Investigating autocatalytic gene expression systems through mechanistic modeling. *J. Theor. Biol.* 201:25–36.
14. Khlebnikov, A., O. Risa, T. Skaug, T. A. Carrier, and J. D. Keasling. 2000. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.* 182:7029–7034.
15. Khlebnikov, A., K. A. Datsenko, T. Skaug, B. L. Wanner, and J. D. Keasling. 2001. Homogeneous expression of the P-BAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Part 12. *Microbiology*. 147:3241–3247.
16. Morgan-Kiss, R. M., C. Wadler, and J. E. J. Cronan. 2002. Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. *Proc. Natl. Acad. Sci. USA*. 99:7373–7377.
17. Mettetal, J. T., D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden. 2006. Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. USA*. 103:7304–7309.
18. Longo, D., and J. Hasty. 2006. Dynamics of single-cell gene expression. *Mol. Syst. Biol.* 2:64.
19. Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P-BAD promoter. *J. Bacteriol.* 177:4121–4130.
20. PROMEGA. 1994. Altered sites II *in vitro* mutagenesis system. *In* Promega Technical Manual. Promega Corporation, Fitchburg, WI.
21. Cormack, B. P., R. H. Valdivia, and S. Falkow. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*. 173:33–38.
22. Andersen, J. B., C. Sternberg, L. K. Poulsen, S. P. Bjorn, M. Givskov, and S. Molin. 1998. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* 64:2240–2246.
23. Sambrock, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
24. Abramoff, M., P. Magelhaes, and S. Ram. 2004. Image processing with ImageJ. *Biophoton. Intl.* 11:36–42.
25. Gordon, A., A. Colman-Lerner, T. E. Chin, K. R. Benjamin, R. C. Yu, and R. Brent. 2007. Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nat. Methods*. 4:175–181.
26. Schleif, R. 1969. Induction of the L-arabinose operon. *J. Mol. Biol.* 46:197–199.
27. Tsien, R. Y. 1998. The green fluorescent protein. *Annu. Rev. Biochem.* 67:509–544.
28. Nagai, T., K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, and A. Miyawaki. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* 20:87–90.
29. Cronan, J. E. 2006. A family of arabinose-inducible *Escherichia coli* expression vectors having pBR322 copy control. *Plasmid*. 55:152–157.
30. Lin-Chao, S., W.-T. Chen, and T.-T. Wong. 1992. High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Mol. Microbiol.* 6:3385–3393.
31. Paulsson, J., and M. Ehrenberg. 2001. Noise in a minimal regulatory network: plasmid copy number control. *Q. Rev. Biophys.* 34:1–59.
32. Cooper, S. 2006. Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theor. Biol. Med. Model.* 3:10.
33. Bernstein, J. A., A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA*. 99:9697–9702.
34. Smolke, C. D., T. A. Carrier, and J. D. Keasling. 2000. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Appl. Environ. Microbiol.* 66: 5399–5405.
35. Nilsson, G., J. G. Belasco, S. N. Cohen, and A. von Gabain. 1984. Growth-rate dependent regulation of mRNA stability in *Escherichia coli*. *Nature*. 312:75–77.
36. Jain, C., and J. G. Belasco. 1995. RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in *Escherichia coli*: unusual sensitivity of the RNA transcript to RNase E activity. *Genes Dev.* 9:84–96.
37. Becskei, A., and L. Serrano. 2000. Engineering stability in gene networks by autoregulation. *Nature*. 405:590–593.
38. Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. 2005. Gene regulation at the single-cell level. *Science*. 307:1962–1965.

39. Daruwalla, K. R., A. T. Paxton, and P. J. Henderson. 1981. Energization of the transport systems for arabinose and comparison with galactose transport in *Escherichia coli*. *Biochem. J.* 200:611–627.

40. Berg, O. G. 1978. A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* 71:587–603.

41. Thattai, M., and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA.* 98:8614–8619.

42. Kolodrubetz, K., and R. Schleif. 1981. Regulation of the L-arabinose transport operons in *Escherichia coli*. *J. Mol. Biol.* 151:215–227.

43. Johnson, C., and R. Schleif. 1995. In vivo induction kinetics of the arabinose promoters in *Escherichia coli*. *J. Bacteriol.* 177:3438–3442.

44. Kepler, T. B., and T. C. Elston. 2001. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–3136.

45. Timmes, A., M. Rodgers, and R. Schleif. 2004. Biochemical and physiological properties of the DNA binding domain of AraC protein. *J. Mol. Biol.* 340:731–738.

46. Neidhardt, F. C., J. L. Ingraham, and M. Schaechter. 1990. Physiology of the Bacterial Cell. Sinauer Associates, Sunderland, MA.

47. Kehres, D. G. 1992. A kinetic model for binding protein-mediated arabinose transport. *Protein Sci.* 1:1661–1665.

48. Kashket, E. R. 1981. Effects of aerobiosis and nitrogen source on the proton motive force in growing *Escherichia coli* and *Klebsiella pneumoniae* cells. *J. Bacteriol.* 146:377–384.

49. Kaback, H. R., and J. Wu. 1997. From membrane to molecule to the third amino acid from the left with a membrane transport protein. *Q. Rev. Biophys.* 30:333–364.

50. Bruinsma, R. F. 2002. Physics of protein-DNA interaction. *Physica A.* 313:211–237.

51. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.

52. Miyada, C. G., L. Stoltzfus, and G. Wilcox. 1984. Regulation of the araC gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on araBAD expression. *Proc. Natl. Acad. Sci. USA.* 81:4120–4124.

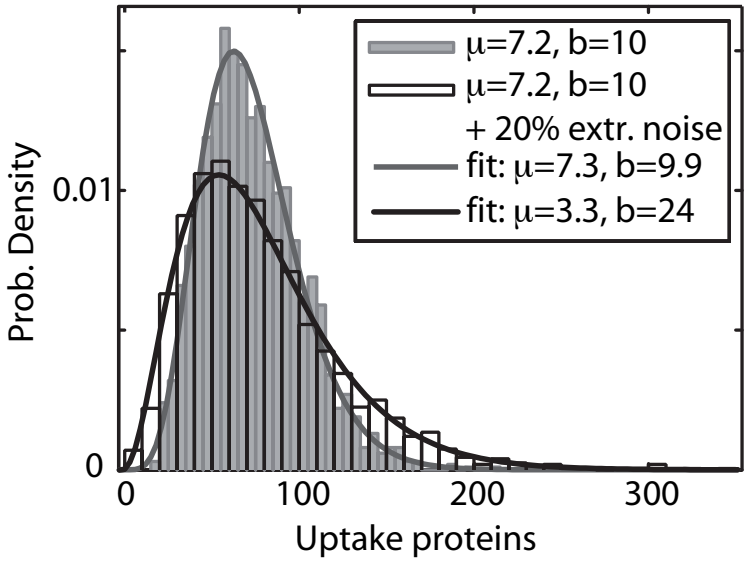


Figure S1: Influence of extrinsic noise on the steady-state distribution of uptake proteins. Stochastic simulations in the absence of extrinsic noise lead to a protein distribution (*grey bars*) that is well described by Eq. 2 in the main paper, and a fit with a negative binomial (*grey line*) recaptures the characteristic parameters μ and b. Addition of a normally distributed source of noise to all parameters (choosing a typical coefficient of variation of 20% (1)), leads to a significantly widened distribution of uptake proteins (*black bars*). The shape of this distribution is still well described by a negative binomial (*black line*), but with a higher burst factor b and a reduced burst frequency μ.

References

1. Rosenfeld, N., J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz, 2005. Gene regulation at the single-cell level. *Science* 307:1962–1965.

Quantitative Characterization of Single Cell Switching Dynamics in the Arabinose Utilization System

Running title: Switching dynamics of the *ara* system

Judith A. Megerle,[†] Georg Fritz,[‡] Sonja A. Westermayer,[†] Delia Brick,[†] Ralf Heermann,[§] Kirsten Jung,[§] Joachim O. Rädler,[†] and Ulrich Gerland[‡]

[†] Faculty of Physics and CeNS, Ludwig-Maximilians-Universität, Munich, Germany

[‡]Arnold Sommerfeld Center for Theoretical Physics and CeNS, Ludwig- Maximilians-Universität, Munich, Germany

[§]Center for Integrated Protein Science (CiPSM) at the Department of Biology, Microbiology, Ludwig-Maximilians-Universität, Martinsried, Germany

Judith A. Megerle and Georg Fritz contributed equally to this work.

Abstract

Adaptation to changing environments often requires rapid switching between phenotypes. Here, we analyze the single-cell transcription dynamics of the arabinose system in several mutants of *E. coli*, using time-lapse fluorescence microscopy and quantitative modeling. While there is significant, inducer-dependent cell-to-cell variation in the timing of the on-switching, the off-switching triggered by sudden removal of arabinose is rapid and homogeneous. We find that rapid off-switching is dependent on the poorly characterized membrane protein AraJ in cells that are unable to metabolize arabinose. AraJ appears to mediate arabinose efflux, thereby quickly reducing the internal arabinose level below the induction threshold after removal of external arabinose. We also find that the cell-to-cell variation in the timing of the on-switching disappears in a mutant in which the positive feedback on the arabinose uptake protein AraE is disrupted. Instead of a switch-like but heterogeneous response, this mutant displays a graded but homogeneous response consistent with our quantitative model. Taken together, our results give new insight into the regulatory dynamics of one of the best-studied systems in bacterial gene regulation.

Keywords: Sugar exporter; time-lapse microscopy; cell-to-cell variation; quantitative model; inducible system; gene regulation

Introduction

Signalling pathways and gene regulatory circuits enable bacteria to respond to environmental changes by turning functional genetic modules on or off. Examples for such modules include conditional stress response systems (1-4), cell-to-cell communication modules (5,6) and inducible carbon utilization systems (7,8). In many cases, noise in gene expression strongly influences the switching process (9,10). Owing to recent technological advances, the gene expression dynamics of functional modules in single cells of bacterial populations can now be monitored in real time (11), while their environments can be controlled using microfluidic devices (12). The resulting experimental portrait of these systems, combined with theoretical analysis, can yield unprecedented insight into the strategies for cellular decision making (13).

The inducible sugar utilization systems of *E. coli* display a clearcut switching behavior and are well suited for quantitative analysis. The lactose (*lac*) and the arabinose (*ara*) system in particular are paradigmatic examples of negative and positive transcription regulation, respectively, and are well characterized at the molecular level 14,15). In both systems, a regime of bimodal distributions in the expression levels of cells within a population ("all-or-nothing gene expression") has been observed (7,8,16) and its molecular determinants have been elucidated (17-19). For the arabinose system, we previously characterized the single-cell induction dynamics using time-lapse fluorescence microscopy and identified a heterogeneous timing behavior, where the time point of gene induction is highly variable between cells of a clonal population (20). This variability depends sensitively on the external arabinose concentration and theoretical analysis of the underlying regulatory circuit revealed that the stochastic distribution of arabinose transporters at the time of sugar addition consistently accounts for these observations. In contrast, the dynamics of switching from the on- to the off-state, triggered by a sudden removal of arabinose, has not been characterized in a similar way. Does the process of switching off display homogeneous or heterogeneous timing? Can we improve our quantitative understanding of the arabinose system by analyzing its dynamics of downregulation?

A study of the downregulation process could in fact help to clarify the role of a functionally poorly characterized component of the *ara* system (Fig. 1), the *araJ* gene (15,21): Since the induction of the *ara* system is coupled to an intracellular sensing of the arabinose level via the key regulator AraC, the transcriptional response upon external sugar removal is expected to critically depend on the import, metabolization, and export dynamics for arabinose. The *ara* system features two transport systems for arabinose uptake (AraE, AraFGH) and the AraBAD operon, which is responsible for arabinose catabolism. However, there is also evidence for an export mechanism, from a mutant strain ($araA^-$) deficient in arabinose catabolism (22). The product of the fourth arabinose-inducible gene, *araJ*, could indeed be involved in arabinose export, since sequence similarity with drug efflux proteins (23) suggests that it may function as a sugar efflux system (24). Thus, analogous to the heterogeneity in the induction behavior, a statistical distribution of arabinose efflux pumps might result in a heterogeneous timing in switching the arabinose system off.

To test these hypotheses, we study the downregulation of the *ara* system in a mutant deficient of arabinose metabolization, using a microfluidic setup where the sugar can be rapidly added and removed from the bacterial environment. In this strain, the downregulation dynamics should be (i) dominated by arabinose export and (ii) sensitive to deletion of *araJ*, if it does indeed provide the primary export pathway. Our results confirm this expectation and, on a quantitative level, lead us to a refined model for the *ara* system. In particular, the model predicts that the arabinose export should also affect the upregulation of the *ara* system in a way that is sensitive to deletion of the positive feed-

back loop of arabinose uptake proteins on their own synthesis (see Fig. 1). We test this prediction experimentally using a mutant strain with an unregulated basal AraE transporter expression. In accordance with the model, we find that this mutation changes the behavior of the *ara* system from a binary into a graded response, where the rate of gene expression is modulated by the external arabinose concentration, similar to the observations previously reported in (25,26). In summary, our results lead us to a quantitative understanding of the single-cell switching dynamics in several mutants of the arabinose system.

Materials and Methods

Bacterial strains and plasmid

In this study *E.coli* strains BW25113 (native reference strain containing AraJ), JW0386-1 (*araJ* deleted), JW1889-3 (native *araE* regulation) and JW1889-5 (*araE* under control of P_{lac}) were used (see Supplemenatry Material for details). In all strains the *araBAD* genes are deleted. All strains were transformed with the reporter plasmid pBAD24-GFP (20), containing the gene encoding *gfpmut3* (27) under control of the P_{BAD} promoter as well as the *araC* gene. Plasmid pCP20 (28) was used for elemination of gene fragments between FTR sites. Plasmid pRed/ET(amp) (GeneBridges, Heidelberg) encodes the lamda Red recombinase system *redαβγ/recT* and was used for Red/ET recombination.

Growth conditions and induction experiments

Cells were grown in LB broth (29) under aerobic conditions at the designated temperature. For solid media, 1.5% (w/w) agar was added. Antibiotics were used at the following final concentrations: ampicillin (50 µg/ml), carbenicillin (50µg/ml in solid media), kanamycine (15µg/ml), and streptomycine (50µg/ml). Overnight pre-cultures were inoculated from single colonies grown on LB agar plates. Cultures were inoculated 1:400 from the precultures and subsequently grown for 3 to 4h (OD600 ~ 0.1-0.2), at which time they were prepared for microscopy. All liquid cultures were grown at 37°C in M63 medium (30) containing 0.2% (v/v) or 0.5% (v/v) glycerol as C-Source. For microscopy, bacteria were applied to one channel of a poly-L-lysine coated microfluidic chamber (µ-slide VI; Ibidi, Martinsried, Germany). The slide was subsequently incubated at 37°C for several minute, rinsed with fresh, pre-warmed medium and transferred to the microscope. For induction experiments the channel was flushed several times with pre-warmed medium containing the desired arabinose concentration after acquiring two images. To generate arabinose pluses arabinose was removed at the indicated time after induction by rinsing the channel several times with pre-warmed medium without arabinose.

Time-lapse microscopy and image analysis

Time-lapse microscopy was performed as described previously (20): Briefly, bright-field and fluorescence images of several fields in one sample were acquired every 5 min at 100x magnification. Image analysis was performed in a semi-automated way using the ImageJ plugin CellEvaluator (31). Using this program, bacterial outlines were determined in the bright-filed images and tracked through the entire time-series. Subsequently the outlines were transferred to the background corrected fluorescence images and the time course of the total fluorescence (sum over all pixel values within the outline) was extracted.

Mathematical model for switching dynamics

A rate equation model for arabinose uptake and gene regulation via internal arabinose was devised. It served us to understand the impact of different factors on the systems dynamics and was also used to analyze the single-cell fluorescence trajectories in Figs. 2 and 5. Comparison of arabinose uptake in wild-type strains with *araE* and *araFGH* deletion strains revealed that the two transporters do not operate independently and that for intermediate to high sugar levels AraE is the dominant uptake protein (32). Our model reflects this behavior through the use of AraE as the single uptake protein species for arabinose import. The temporal behavior of internal arabinose, a(t), is then determined by an interplay of sugar import at rate V_{upt} by AraE, E(t), sugar dilution by cell growth at rate γ and arabinose export by a yet unknown mechanism at rate k (22):

$$\partial_t a(t) = V_{upt} \times E(t) - (k + \gamma) \times a(t), \qquad (1)$$

Here, based on experimental findings of Novotny and Englesberg (22), sugar export follows first-order reaction kinetics, while the rate of sugar uptake depends on the external sugar concentration, a_{ex}: $V_{upt} = V_{max} \times a_{ex} / (K_m + a_{ex})$, where K_m is the Michealis-Menten constant and V_{max} the maximal arabinose uptake velocity per AraE protein. In the presence of sufficient amounts of internal arabinose, AraC stimulates transcription from P_E and P_{BAD} and the promoter activity of P_{BAD} increases cubically with the internal arabinose concentration (33). While the quantitative characteristics of P_E are not known so far, the high similarity between P_E and P_{BAD} suggests that the input-output relations of both promoters are similar (34). Hence, we model transcriptional regulation in both cases as cubic Hill functions,

$$\upsilon = \upsilon_0 + (\upsilon_{max} - \upsilon_0) \times \frac{a^3}{K_a^3 + a^3}, \qquad (2)$$

where ν_0 and ν_{max} are the basal and maximal transcription rates of each promoter, respectively. Note that we do not consider the dynamics of AraC explicitly, since AraC negatively regulates its own expression (15), resulting in homeostatic control of its number. With this the dynamics of AraE mRNA, e(t), and GFP mRNA, g(t), reads

$$\partial_t e(t) = \upsilon_e - (\lambda_e + \gamma) \times e(t), \text{ and} \qquad (3)$$

$$\partial_t g(t) = \upsilon_g - (\lambda_g + \gamma) \times g(t), \qquad (4)$$

where λ_e and λ_g are the degradation rates of AraE and GFP mRNA, respectively. Likewise, protein dynamics of AraE, E(t), immature GFP, I(t), and fluorescent GFP, G(t), obey the rate equations

$$\partial_t E(t) = \mu_e \times e(t) - \gamma \times E(t), \qquad (5)$$

$$\partial_t I(t) = \mu_g \times g(t) - (\tau_m^{-1} + \gamma) \times I(t), \text{ and} \qquad (6)$$

$$\partial_t G(t) = \tau_m^{-1} \times I(t) - \gamma \times G(t). \qquad (7)$$

Here τ_m^{-1} is the time constant of GFP maturation, the process whereby the folded GFP becomes fluorescent. Finally, for a direct comparison of model and experiment we need to con-

sider the total fluorescence per cell, $F(t) = \sigma \times e^{\gamma t} \times G(t)$ which accounts for the fluorescence increase by exponential cell growth and converts protein numbers into arbitrary fluorescence units via a scaling factor σ. The two key parameters, $\nu_{0,e}$ and $\nu_{max,g}$, which are expected to dominate the cell-to-cell variability in the experimental response, were estimated by solving the reaction system in Eqs. (1)-(7) numerically and by using a trust-region reflective Newton method (MATLAB, The MathWorks, Inc.) to minimize the total χ^2 between experiment and model

Results

Single cell switching from the on- to the off-state

To study the switching from the on- to the off-state in the *ara* system, we use an *E. coli* strain transformed with the reporter plasmid pBAD24-GFP (20) that contains the *araC* gene and the rapidly maturing GFP variant *gfpmut3* (27) under the control of the P_{BAD} promoter. A chromosomal deletion of *araBAD* avoids the negative feedback of arabinose metabolization and thus allows us to isolate the effect of arabinose efflux on the switching dynamics. The *araC* gene is supplied on the plasmid to guarantee full functionality of the DNA loop required for repression of P_{BAD} in the absence of arabinose (15) and to provide the proper stoichiometry of transcription factors and P_{BAD} promoters.

To prepare cells in the on-state, we first flush the microfluidic channel with medium containing saturating amounts of arabinose (0.2% ara) at t = 0 min. After 20 min, we remove the external stimulus by flushing with fresh, arabinose-free medium. The resulting single-cell dynamics are shown in Fig. 2A. After the initial increase of total fluorescence during the induction phase (*grey area*), fluorescence trajectories reach a plateau about 20-30 min after arabinose removal. Since GFP is a stable protein, this indicates that all cells switch off the arabinose system after arabinose removal.

In order to extract the switching dynamics at the transcriptional level, and to probe our quantitative understanding of the system, we devise a mathematical model for the single-cell expression dynamics in the *ara* system. The smooth shape of the time series suggests that the dynamics of individual cells follows a rather deterministic fate, while the differences between the cells stem from cell-to-cell variation of the reaction rates. Therefore, we use a deterministic rate equation model to describe the single-cell dynamics of arabinose uptake, export and gene regulation by internal arabinose, but allow for cell-to-cell variations in the model parameters. This model refines our previous approach in which the induction process of the *ara* system was simply described by a step-like increase of the promoter activity after a time delay τ_d (20). We now explicitly include the experimentally measured dependence of the transcription rate on the internal arabinose concentration (33), see Materials and Methods for all details.

We use the same protocol of pulsed arabinose addition as in our experiments and fit the model to the single-cell induction kinetics in Fig. 2A. To that end we fix most of the model parameters to values extracted or estimated from the literature or our own measurements (see Table 1) and vary only two key parameters which are expected to display strongest cell-to-cell variation (20): (1) the basal transcription rate of *araE* (encoding the transporter AraE) from the P_E promoter, $\nu_{0,e}$, and (2) the maximal expression rate of GFP, $\nu_{max,g}$. Note that this choice fixes all dynamical timescales in our model and only allows for a variation of the onset (via $\nu_{0,e}$) and the amplitude (via $\nu_{max,g}$) of the response.

Fig. 2A highlights a few representative single-cell trajectories (*circles*) and shows their cognate fits with our quantitative model (*bold lines*). Fig. 2B shows the corresponding dy-

namics of P_{BAD} promoter activities in individual cells as extracted from the model. During the induction phase (t=0-20 min), promoters generally switch on within ~10 min from low to high values, but there is a certain variation between the induction times of individual cells. When we use the time at which the promoter activity reaches 95% of its maximal level as a proxy for this time delay τ_d, we find a distribution of delay times that is consistent with our previous experiments and model (20), as shown in Fig. S2 in the Supporting Material.

As shown in Fig. 2B, the data of Fig. 2A is consistent with a promoter activity that drops almost instantaneously to basal levels at the end of the arabinose pulse in all cells of the culture. According to the model, the increase of fluorescence after removal of arabinose is caused by the translation of existing GFP mRNA molecules prior to their decay and the maturation of residual, non-fluorescent GFP molecules. Note that we have not fitted the arabinose export rate, but taken the value determined in (22), see Table 1. Due to the relatively long (~7 min) maturation timescale of GFP, we cannot infer the export rate from the data. However, Fig. 2 shows that our experiments are compatible with the fast export rate of Ref. (22).

Taken together, these results confirm heterogeneous timing in the switching-on dynamics of the arabinose system, and reveal that there is no such heterogeneity in the timing of the switching-off process. Within our model, this absence of heterogeneity is a direct consequence of the fast export rate.

Effect of AraJ on switching dynamics

So far, the molecular mechanism for rapid loss of internal arabinose is unclear, given that the *araBAD* operon required for arabinose metabolization is not present in our strain. However, it has long been known that besides *araBAD*, *araE*, and *araFGH*, there is also a fourth arabinose-inducible gene in *E. coli*, *araJ* (35). While there is an apparent lack of function of AraJ for the induction of the *araBAD* operon (21), sequence similarity with drug efflux proteins of the major facilitator superfamily (23) suggests that it may function as an efflux system for arabinose (24). To test whether AraJ has an effect on the dynamics of the switching-off process, we compare the response of a Δ*araJ* mutant to that of the corresponding strain with the native *araJ* gene, when both strains are exposed to an arabinose pulse as above. Fig. 3 strikingly shows that the Δ*araJ* mutant displays an increasing fluorescence for more than 50 min after arabinose removal, in sharp contrast to the strain with *araJ*. This constitutes a clear phenotype for AraJ, which is observable only in the downregulation of the arabinose system.

A possible explanation for the behavior shown in Fig. 3 is that AraJ acts indeed as an efflux system for arabinose, and that the lack (or strong reduction) of arabinose efflux in the Δ*araJ* mutant prevents a rapid equilibration of the internal arabinose upon the removal of external arabinose. It is then plausible that the P_{BAD} promoter activity is downregulated on a much slower timescale, e.g. the timescale of cell growth, which effectively dilutes the internal arabinose.

Influence of arabinose efflux on induction kinetics

Since the analysis of the downregulation dynamics led us to extend our previous model (20), we next study the effect of the added term for sugar export on the induction kinetics, in order to test whether the refined model yields a coherent description of the arabinose system. In addition, to further validate the model, we use it to predict the induction kinetics of a mutant lacking the positive autoregulation of the arabinose transporters.

To that end, we first calculate the model dynamics upon a sudden addition of arabinose, using the physiological parameters of the *ara* system (Table 1). We consider three different external arabinose concentrations (leading to different uptake velocities V_{upt} per transporter). To illustrate the effect of the cell-to-cell variation in the number of uptake proteins, we show the model dynamics at each arabinose level for four representative initial uptake protein numbers (300, 260, 220 and 180 AraE proteins). As shown in Fig. 4A, all trajectories increase shortly after adding a high level of arabinose, whereas at a low arabinose level the model reproduces the heterogeneous time delay until full induction of GFP expression. Thus, the model is qualitatively consistent with the experimental induction behavior (20).

Within our refined model, the detailed mechanism behind this delay is as follows: Upon arabinose addition, the internal arabinose level rapidly equilibrates to a steady state value a_{in}^* determined by the ratio of arabinose in- and efflux rates, i.e., $a_{in}^* = \left(E(t) \times V_{upt}\right)/k$, where E(t) is the number of uptake proteins at time t and k is the rate of arabinose efflux. The cell-to-cell variability in the initial number of arabinose uptake proteins then leads to a distribution of a_{in}^* values. Since internal arabinose activates the *ara* genes according to Eq. (2), this leads to a cell-to-cell variability in the initial rates of gene expression (Fig. 4B). Due to the positive feedback of the uptake proteins on their own expression, the cellular level of transporters increases, while the internal arabinose concentration follows adiabatically, that is, immediately after a change in transporter level. As a result, the promoter activity increases on a timescale determined by the (slow) kinetics of *araE* expression (Fig. 4B).

Due to the variation in the initial steady state value a_{in}^*, the maximal value of the promoter activity is reached earlier in cells with a high initial number of uptake proteins than in cells with a low initial number of uptake proteins. The delay time until promoter activation in the refined model displays approximately the same $\tau_d \sim 1/E$ dependence as in the previous model over the range of typical basal transporter expression levels [~20-200 proteins (36)], see Fig. S1. Furthermore, the distribution of the delay times has almost the same shape in the two models, see Fig. S2, such that both models are also quantitatively consistent with the experimental switching kinetics from the off- to the on-state of the *ara* system.

Interestingly, an important difference between the two models arises when the positive auto-regulation of AraE expression is disrupted: The refined model predicts that a mutant with constitutive AraE expression will no longer exhibit distinct switching to full induction after a well-defined time delay (Fig. 4C). Instead, the promoter activity quickly reaches a constant value according to the internal steady-state value of arabinose (Fig. 4D), which in turn determines the slope of the fluorescence trajectory. Thus, the model with arabinose efflux predicts that in such a mutant fluorescence levels will rise almost instantly with a sub-maximal slope. In contrast, in the model without arabinose efflux the induction kinetics is virtually unaffected by the positive feedback (20), and fluorescence trajectories are expected to display delayed induction both in the strains with and without positive feedback, see Fig. S3. This is due to the fact that in a model without sugar efflux the experimentally observed delay times are explained by slow arabinose uptake, such that the internal arabinose concentration initially increases linearly at a rate determined by the initial number of transporters. The delay time is then largely independent of the positive feedback, since it is mainly caused by slow import of arabinose rather than by slow protein dynamics involved in feedback activation.

Single-cell switching from the off- to the on-state

To test whether a mutant with disrupted positive feedback displays a graded response as predicted by our model, we constructed a strain where *araE* is expressed at a basal level, in-

dependent of arabinose, in a *araBAD*$^-$/ *araFGH*$^-$ background. The construct is based on the *lac* promoter, which is repressed to a constant low level by the *lac* repressor; see Supporting Text for details. As a reference strain, we used the same genetic background, but with native, P_E-driven expression of *araE* (Fig. 5; *top panels*). Fluorescence trajectories of the reference strain are shown in Fig. 5A-C for 0.2%, 0.05% and 0.01% arabinose, corresponding to 13.3 mM, 3.33 mM and 0.66 mM arabinose, respectively. The strain with constitutive *araE* expression was sensitive already at lower arabinose levels, consistent with the fact that the repression of our constructed promoter is relatively leaky. The corresponding fluorescence trajectories are shown for 0.2% (13.3 mM), 0.002% (0.133 mM) and 0.001% (0.066 mM) arabinose, see Fig. 5D-F.

A qualitative comparison of the single-cell traces shows that the dynamics of the two strains differ significantly. In the case of feedback-regulated *araE* expression (Fig. 5A-C), cells display a distinct lag-phase with marginal fluorescence increase and an induction-phase with high rate of fluorescence production. As the external arabinose concentration decreases, the average as well as the cell-to-cell variation of the delay time increases (Fig. 5A-C). This is qualitatively consistent with the previously described behavior in strain LMG194 (20), an *araBAD*$^-$ strain in which both *araE* and *araFGH* are functional and under positive feedback control. In contrast, while in the strain with constitutive *araE* expression fluorescence increase starts immediately after arabinose addition at all tested concentrations (Fig. 5D-F), the rate of fluorescence production decreases markedly with decreasing external arabinose.

To quantify the difference between the two strains, we use the analytical gene expression function of (20), which accounts for a step-like increase of the protein expression rate from zero to σ_p after a delay time τ_d, and determine both parameters by fitting each single-cell trace, see Table2 and Materials and Methods. As expected, the reference strain displays an increasing mean and width of the delay time distribution upon decreasing the arabinose concentration. Meanwhile, the protein expression rates remain virtually unchanged. Notably, the mean values of the delay times obtained here coincide with the previously measured values in *E. coli* strain LMG194 within experimental errors (20), indicating that the transport system AraFGH (which is absent in the reference strain used here) has no significant impact on the delay time distribution over the range of arabinose concentrations used here. This is plausible, since AraFGH constitutes a high affinity/low capacity transporter (32), suggesting that it plays a minor role for the rapid arabinose import at intermediate sugar levels.

For the strain with constitutive araE expression (Fig. 3D-F) we find that the average delay times are small ($<\tau_d>$ =0.1-2.2 min) and do not vary systematically with the external arabinose concentration. In contrast, the average protein expression rate decreases with decreasing arabinose concentration from ~3100 a.u./min at 0.2% arabinose to ~400 a.u./min at 0.001% arabinose (Table 2). Interestingly, even at saturating arabinose levels (0.2% arabinose) the average protein production rate in the strain without positive feedback is only 60% of the production rate in the reference strain. This observation is consistent with the dynamics in the model without positive feedback (Fig. 4D), for which the promoter activity remains submaximal even for very high arabinose uptake velocities.

Finally, we test whether the presented model for the induction kinetics of the *ara* system coherently describes the single-cell dynamics of both strains also on a quantitative level. To that end we again fit the model to the single-cell induction kinetics in Fig. 5 by varying (1) the basal transcription rate of P_E, $\nu_{0,e}$, (reference strain) or P_{lac}, $\nu_{0,lac}$, (mutant strain) and (2) the maximal expression rate of *gfp*, $\nu_{max,g}$,. To account for the different external arabinose concentrations in our experiments, we scale the effective arabinose uptake velocity, V_{upt}, according to Michaelis-Menten kinetics as described in the supporting text.

For both strains and all concentrations tested we find good agreement between model and experiment, as illustrated by a few representative experimental trajectories and their cognate

fits (Fig. 5; *dots* and *bold lines*). For the reference strain, the resulting distributions of fit parameters indicate that all single-cell trajectories are compatible with a single underlying AraE uptake protein distribution at the time of sugar addition (Figs. S4 and S5). Furthermore, they suggest that the Michaelis-Menten scaling of the arabinose uptake velocity with the external arabinose concentration consistently explains the observed modulation of the time delay. For the mutant with constitutive *araE* expression the resulting fit parameters display a strong anti-correlation, suggesting that only the product of $\nu_{0,lac}$ and $\nu_{max,g}$ is well constrained by our data (Figs. S6 and S7). This is expected, since both parameters are proportional to the slope of the fluorescence trajectories in this mutant. We find that the product of both parameters does not vary systematically with the external arabinose concentration (Fig. S8), indicating that in the case of constitutive *araE* expression the Michaelis-Menten scaling of the arabinose uptake velocity alone causes the observed modulation of the protein expression rate.

Discussion

In this paper we provide a detailed experimental and theoretical study of the single-cell switching dynamics in the arabinose utilization system. In addition to traditional induction kinetics, in which cells are exposed to a step-like increase of the inducer, we also measure the dynamics after sudden removal of arabinose. All of our data, for the `off-to-on' and `on-to-off' transitions in several mutants, is consistent with a theoretical model that explicitly incorporates arabinose efflux. Moreover, we show that deletion of AraJ, a previously uncharacterized component of the arabinose system, has a distinct phenotype in the on-to-off transition for a strain that cannot metabolize arabinose: Deletion of AraJ slows down the transition from the on- to the off-state significantly, suggesting that it is involved in arabinose export. From our data, we cannot exclude that there are additional routes of arabinose efflux, but, if so, they act on a longer timescale than AraJ, as seen by the slow decrease of fluorescence production in the Δ*araJ* mutant (Fig. 3; *solid curves*). For instance, arabinose might also exit the cells via the arabinose transporters AraE and AraFGH or via unspecific sugar exporters (37).

While efflux systems are well known to confer resistance to a variety of antimicrobial agents, the physiological advantage of export systems for nutrients is unclear. Toxic effects of high concentrations of sugars may have provided evolutionary pressure for the acquisition of such efflux systems (38). Growth inhibition may also result from the uptake of toxic sugar analogues that inhibit uptake of bona fide sugar substrates or interfere with normal metabolism (37). For instance, the non-metabolizable lactose analogue thiomethylgalactoside (TMG) is exported from *S. pyogenes* (39) and appears to be excreted from E. coli cells as well, since rapid switching from the on- to the off-state is observed in the lac operon when suddenly deprived of external TMG (17). Moreover, upon addition of lactose to a strain with constitutively expressed lac operon, intermediates of lactose catabolism (glucose, galactose, and allolactose) rapidly accumulate in the medium (40) and it was suggested that sugar efflux is an integral part of the metabolism of lactose in *E. coli* (41).

In case of the arabinose system, however, a similar significance for arabinose catabolism seems unlikely, since insertion/deletion mutations in *araJ* had no detectable effect on the ability of the bacteria to grow on arabinose (21). Instead, AraJ may primarily function as efflux valve for arabinose analogues, such as D-xylose and D-fucose: While these compounds are competitive inhibitors for L-arabinose uptake (22) and are thus likely to be imported by the arabinose permease system, their rate of metabolization via the arabinose isomerase AraA is greatly reduced (42). Hence, in the absence of suitable export mechanisms such compounds might accumulate to toxic amounts. At the same time, in order to prevent futile cycling of arabinose in and out of the cell, the rate of arabinose efflux should not exceed the rate of in-

ternal arabinose metabolization. This idea is indirectly supported by the fact that arabinose export follows effective first order reaction kinetics up to millimolar sugar concentrations (22), suggesting that AraJ has a low apparent affinity for arabinose. In the future, a biochemical characterization of AraJ will help to better understand its physiological and regulatory role in the *ara* system. While our results indicate that AraJ is an arabinose exporter, we cannot draw reliable conclusions on the significance of its effect in the native system where arabinose is metabolized. In our simplified genetic background, which is useful for the quantitative characterization of the system, the intracellular arabinose concentration may rise to very high levels. Therefore, the expression of AraJ may be induced to higher levels than in the native situation. Thus, in addition to a biochemical characterization of AraJ, further gene expression studies of the native *ara* system are necessary to clarify the physiological role of AraJ.

We have also shown that arabinose efflux impacts the upregulation dynamics of the P_{BAD} promoter in strains with constitutive arabinose transporter production. Such constructs are often used as gene expression systems (43), e.g., for biotechnological applications, when a controlled, homogeneous expression of proteins in all cells of a culture is desired (25,26,44). Here we found that the system responds to external inducer addition at a rate proportional to the internal sugar level. Thereby, variation of external inducer concentration leads to a direct modulation of the average protein expression rate (Figs. 4 and 5). Our model provides a coherent and mechanistic explanation for the graded, dose-dependent response found in strains with constitutive *araE* expression (25,26). Taken together, our results shed new light on the regulatory dynamics of one of the best-studied systems in bacterial gene regulation and are useful for the redesign of cells in synthetic biology and biotechnological applications.

Acknowledgments

This work was supported by the Excellence Cluster “Nanosystems Initiative Munich” and by the Deutsche Forschungsgemeinschaft via grants RA655/6-1 and GE1098/4-1.

Supporting Material

Supporting text, figures and references (47-52) appear in the Supporting Material.

References

(1) Klipp, E., B. Nordlander, R. Kruger, P. Gennemark, and S. Hohmann. 2005. Integrative model of the response of yeast to osmotic shock. Nat. Biotech. 23:975–982.

(2) Rietkötter, E., D. Hoyer, and T. Mascher. 2008. Bacitracin sensing in *Bacillus subtilis*. Mol. Microbiol. 68:768–785.

(3) Foster, J. 2004. *Escherichia coli* acid resistance: tales of and amateur acidophile. Nat. Rev. Microbiol. 2:898–907.

(4) Fritz, G., C. Koller, K. Burdack, L. Tetsch, I. Haneburger, K. Jung, U. Gerland. 2009. Induction kinetics of a conditional pH stress response system in *Escherichia coli*. J. Mol. Biol. 393:272–286.

(5) Anetzberger, C., T. Pirch, and K. Jung. 2009. Heterogeneity in quorum sensing-regulated bioluminescence of *Vibrio harveyi*. Mol. Microbiol. 73:267–277.

(6) Ng, W. L., and B.L. Bassler. 2009. Bacterial quorum-sensing network architectures. Annu. Rev. Genet. 43:197–222.

(7) Novick, A., and M.Weiner. 1957. Enzyme induction as an all-or-none phenomenon. Proc. Natl. Acad. Sci. USA. 43:553-566.

(8) Siegele, D. A., and J. C. Hu. 1997. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. Proc. Natl. Acad. Sci. USA. 94:8168-8172.

(9) Raj, A., and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135:216-226.

(10) Eldar, A. and M.B. Elowitz. 2010. Functional roles for noise in genetic circuits. Nature 467:167-173.

(11) Locke, J. C. W., and M.B. Elowitz. 2009. Using movies to analyse gene circuit dynamics in single cells. Nat. Rev. Microbiol. 7:383-392.

(12) Bennett, M. R. and J. Hasty. 2009. Microfluidic devices for measuring gene network dynamics in single cells. Nat. Rev. Genet. 10:628-638.

(13) Perkins, T. J., and P. S. Swain. 2009. Strategies for cellular decision-making. Mol. Syst. Biol. 5:326.

(14) Müller-Hill, B. 1996. The Lac Operon: A Short History of a Genetic Paradigm. Walter de Gruyter Inc., Berlin, Germany.

(15) Schleif, R. 2000. Regulation of the L-arabinose operon of *Escherichia coli*. Trends Genet. 16:559-565.

(16) Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden. 2004. Multistability in the lactose utilization network of *Escherichia coli*. Nature 427:737-740.

(17) Mettetal, J. T., D. Muzzey, J.M. Pedraza, E.M. Ozbudak, and A. van Oudenaarden. 2006. Predicting stochastic gene expression dynamics in single cells. Proc. Natl. Acad. Sci. USA. 103:7304-7309.

(18) Choi, P. J., L. Cai, K. Frieda, and X.S. Xie. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science 322:442-446.

(19) Robert, L., G. Paul, Y. Chen, F. Taddei, D. Baigl, and A. B Lindner. 2010. Predispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch. Mol. Syst. Biol. 6:357.

(20) Megerle, J. A., G. Fritz, U. Gerland, K. Jung, and J.O. Rädler. 2008. Timing and dynamics of single cell gene expression in the arabinose utilization system. Biophys. J. 95:2103-2115.

(21) Reeder, T., and R. Schleif. 1991. Mapping, sequence, and apparent lack of function of *araJ*, a gene of the *Escherichia coli* arabinose regulon. J. Bacteriol. 173:7765-7771.
(22) Novotny, C. P., and E. Englesberg. 1966. The L-arabinose permease system in *Escherichia coli* B/r. Biochim. Biophys. Acta 117:217-230.
(23) Pao, S. S., I. T. Paulsen, and M. H. Saier. 1998. Major facilitator superfamily. Microbiol. Mol. Biol. Rev. 62:1-34.
(24) Keseler, I. M., J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muniz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp. 2011. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res. 39:D583-D590.
(25) Khlebnikov, A., K. A. Datsenko, T. Skaug, B. L. Wanner, and J. D. Keasling. 2001. Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology. 147:3241-3247.
(26) Khlebnikov, A., O. Risa, T. Skaug, T. A. Carrier, and J. D. Keasling. 2000. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. J. Bacteriol. 182:7029-7034.
(27) Cormack, B. P., R. H. Valdivia, and S. Falkow. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 173:33–38.
(28) Cherepanov, P. P., Wackernagel, W. 1995. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158:9-14.
(29) Sambrock, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
(30) Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P-BAD promoter. J. Bacteriol. 177:4121–4130.
(31) Youssef, S., S. Gude, and J. O. Rädler. 2011. Automated tracking in live-cell time-lapse movies, Integr. Biol., 3:1095-1101
(32) Daruwalla, K. R., A.T. Paxton, and P.J. Henderson. 1981. Energization of the transport systems for arabinose and comparison with galactose transport in *Escherichia coli*. Biochem. J. 200:611-627.
(33) Schleif, R. 1969. Induction of the L-arabinose operon. J. Mol. Biol. 46:197-199.
(34) Johnson, C. M., and R. F. Schleif. 1995. In vivo induction kinetics of the arabinose promoters in *Escherichia coli*. J. Bacteriol. 177:3438-3442.
(35) Hendrickson, W., C. Stoner, and R. Schleif. 1990. Characterization of the *Escherichia coli* araFGH and araJ promoters. J. Mol. Biol. 215:497–510.
(36) Choi, P.J., L. Cai, K. Frieda, and X. S. Xie. 2008. A Stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science 322:442–446.
(37) Liu, J. Y., P. F. Miller, J. Willard, and E. R. Olson. 1999. Functional and biochemical characterization of *Escherichia coli* sugar efflux transporters. J. Biol. Chem. 274:22977-22984.
(38) Jack, D. L., N. M. Yang, and M. H. Saier. 2001. The drug/metabolite transporter superfamily. Eur. J. Biochem. 268:3620–3639.
(39) Reizer, J., M. J. Novotny, C. Panos, and M.H. Saier. 1983. Mechanism of inducer expulsion in *Streptococcus pyogenes*: a two-step process activated by ATP. J. Bacteriol. 156:354–361.
(40) Huber, R. E., J. Lytton, and E. B. Fung. 1980. Efflux of beta-galactosidase products from *Escherichia coli*. J. Bacteriol. 141:528–533.

(41) Huber, R. E., and K. L. Hurlburt. 1984. *Escherichia coli* growth on lactose requires cycling of beta-galactosidase into the medium. Can. J. Microbiol. 30:411–415.
(42) Patrick, J. W., and N. Lee, 1968. Purification and properties of an L-arabinose isomerase from *Escherichia coli*. J. Biol. Chem. 243:4312–4318.
(43) Keasling, J. D. 2008. Synthetic biology for synthetic chemistry. ACS Chem. Biol. 3:64–76.
(44) Morgan-Kiss, R., C. Wadler, and J. Cronan. 2002. Long-term and homogeneous regulation of the *Escherichia coli* araBAD promoter by use of a lactose transporter of relaxed specificity. Proc. Natl. Acad. Sci USA. 99:7373–7377.
(45) Smolke, C. D., T. A. Carrier, and J. D. Keasling. 2000. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. Appl. Environ. Microbiol. 66:5399-5405.
(46) Thattai, M., and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. USA. 98:8614–8619.
(47) Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 2006.0008.
(48) Datsenko, K. A. and B. L. Wanner. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA. 97:6640-6645.
(49) Heermann, R., T. Zeppenfeld, K. Jung. 2008. Simple generation of site-directed point mutations in the Escherichia coli chromosome using Red(R)/ET(R) Recombination. Microb. Cell Fact. 7:14.
(50) Casadaban, M. J. 1976. Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. J. Mol. Biol. 104:541-555.
(51) Wright, J. K., and P. Overath. 1984. Purification of the lactose: H+ carrier of *Escherichia coli* and characterization of galactoside binding and transport.' Eur. J. Biochem. 138:497–508.
(52) Dornmair, K., P. Overath, and F. Jä□hnig. 1989. Fast measurement of galactoside transport by lactose permease. J. Biol. Chem. 264:342–346.

Parameter	Notation	Value	Source
Arabinose uptake velocity	V_{max}	2000 ara molecules/protein/min	estimated in (20)
Michaelis-Menten constant for arabinose uptake via AraE	K_m	0.3 mM	fitted within narrow, physiological range (32), see Supporting text for details
Arabinose export rate	k	5.2 min^{-1}	(22)
Arabinose binding threshold	K_a	50 μM	(33)
Maximal transcription rate of P_E	$\nu_{max,e}$	4.2 min	inferred from (34)
Basal transcription rate of P_E	$\nu_{0,e}$	fitted*	this study
Basal transcription rate of P_{lac}	$\nu_{0,lac}$	fitted*	this study
Maximal transcription rate of P_{BAD}	$\nu_{max,g}$	fitted*	this study
AraE mRNA degradation rate	λ_e	0.347 min^{-1}	(34)
GFP mRNA degradation rate	λ_g	0.116 min^{-1}	(45)
AraE translation rate	μ_e	10.4 min^{-1}	unknown; chosen to meet typical burst size of 30 (46)
GFP translation rate	μ_g	3.5 min^{-1}	unknown; chosen to meet typical burst size of 30 (46)
Cell doubling time	$\ln(2)/\gamma$	50 min	this study
GFP maturation time	τ_m^{-1}	6.5 min	(20)
Scaling between #GFP and total fluorescence	σ	50 FU/protein	arbitrary

Table 1:
Parameters used in the mathematical model.

* Transcription rates were estimated from fits to single-cell fluorescence trajectories in Figs. 2 and 5 and both their distributions and cross correlations are shown in Figs. S4-S8.

arabinose	$\langle\tau_d\rangle$ [min]	$\sigma(\tau_d)$ [min]	$\langle\alpha_p\rangle$ [FU/min]	$\sigma(\alpha_p)$ [FU/min]
P_E-*araE*				
0.2%	2.4	2.0	4924	1871
0.05%	9.8	5.1	6062	2224
0.01%	27.0	12.9	5390	2056
P_{lac}-*araE*				
0.2%	2.2	1.5	3107	1149
0.02%	0.8	1.0	1520	843
0.002%	1.5	1.8	997	578
0.001%	0.1	0.5	399	351

Table 2:
Mean and standard deviation of delay time τ_d and protein expression rate α_p distribution.

Figure titles and legends

Figure 1:
Scheme of the arabinose utilization system in *E. coli* (Schleif 2000). Arabinose is imported via the arabinose transporters AraE and AraFGH. If the intracellular arabinose level is sufficiently large, arabinose binds the transcriptional regulator AraC. This complex activates the promoters P_E, P_{FGH}, P_{BAD} and P_J, driving expression of *araE*, *araFGH*, *araBAD* and *araJ*, respectively. The latter two operons encode proteins for arabinose catabolism (AraBAD) and a putative arabinose efflux pump (AraJ). The negative autoregulation of AraC resulting in homeostatic control of its own level is not shown explicitly. Arrows indicate arabinose transport and positive regulation, whereas the T-shaped arrow indicates arabinose metabolization.

Figure 2:
Single-cell switching dynamics from the on- to the off-state of the *ara* system. Initially, cells of strain BW25113 (deletion of *araBAD* , otherwise native *ara* regulation) containing the reporter plasmid pBAD24-GFP (*gfpmut3* under the control of the P_{BAD} promoter) were prepared in the on-state by induction with 0.2% arabinose for 20 min. The transition to the off-state was triggered by flushing the microfluidic channel with fresh medium without arabinose. (A) Grey lines are experimental fluorescence trajectories, the symbols highlight representative cells and the bold lines correspond to their cognate fits of the model dynamics [Eqs. (1)-(7)] under pulsed addition of arabinose. FU: fluorescence units. (B) The corresponding dynamics of P_{BAD} promoter activities in individual cells as inferred from the model.

Figure 3:
Single cell gene expression kinetics of a Δ*araJ* mutant (JW0386-1; solid lines) in response to a 20 min pulse of arabinose. For better comparison the data of the reference strain containing AraJ (BW25113; Fig. 2) are replotted (dashed lines) and all trajectories are normalized to their fluorescence levels at the time when arabinose was removed.

Figure 4: Theoretical gene expression kinetics for (A,B) native transporter regulation and (C,D) constitutive transporter expression. Different external arabinose concentrations are represented by various arabinose uptake velocities V_{upt}. To illustrate stochastic variability in the initial number of uptake proteins, trajectories are shown for 300, 260, 220 and 180 AraE molecules at each V_{upt}. In all cases, the corresponding number of initial AraE molecules of the shown curves decreases from left to right. (A) In case of native transporter regulation, trajectories display a homogeneous and rapid response at high V_{upt}, whereas at low V_{upt} there is a significant delay in the response and substantial heterogeneity between the trajectories. (B) Dynamics of the normalized promoter activity under the same parameters as in (A). Initially, the promoter activity rises rapidly to a value determined by the internal arabinose steady-state level, a_{in}^*, which in turn depends on V_{upt} and the number of transporters per cell. Accordingly, transporters are produced and as their level rises, also a_{in}^* and thus the promoter activity follow almost instantaneously. Hence, the full activation of the promoter is reached after a delay incurred by the slow kinetics of transporter expression. (C) In case of constitutive *araE* expression, fluorescence production initiates rapidly, but with sub-maximal rate for all values of V_{upt}. Here the rate of fluorescence production again depends on a_{in}^*, but due to the lack of positive autoregulation, no further transporters are produced and a_{in}^* remains constant (D). Thus, the lack of the positive feedback causes a modulation of the

average expression rate rather than a modulation of the average time delay, as in the case of native transporter regulation.

Figure 5: Single-cell induction kinetics of cells with native (A-C) and constitutive (D-F) transporter production at different arabinose concentrations. Grey lines represent the kinetics of total fluorescence in individual cells and dots highlight representative trajectories. Note that some trajectories disappear before the end of the experiment due to detachment of daughter cells after cell division. Bold lines are representative fits of our full model of arabinose uptake and gene regulation in Eqs. (1)-(7) to the highlighted experimental trajectories.

Figure 1

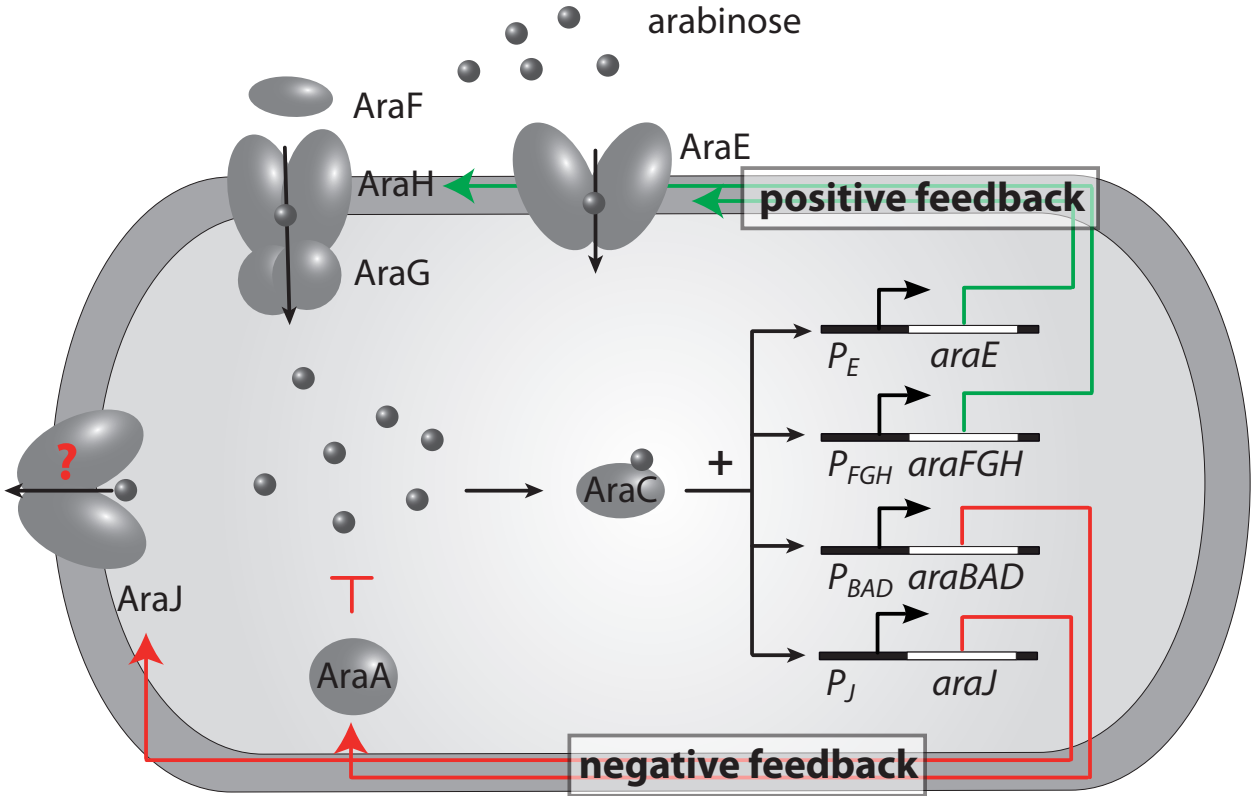

arabinose
AraF
AraE
AraH
positive feedback
AraG
P_E araE
?
AraC
+
P_{FGH} araFGH
P_{BAD} araBAD
AraJ
AraA
P_J araJ
negative feedback

Figure 2

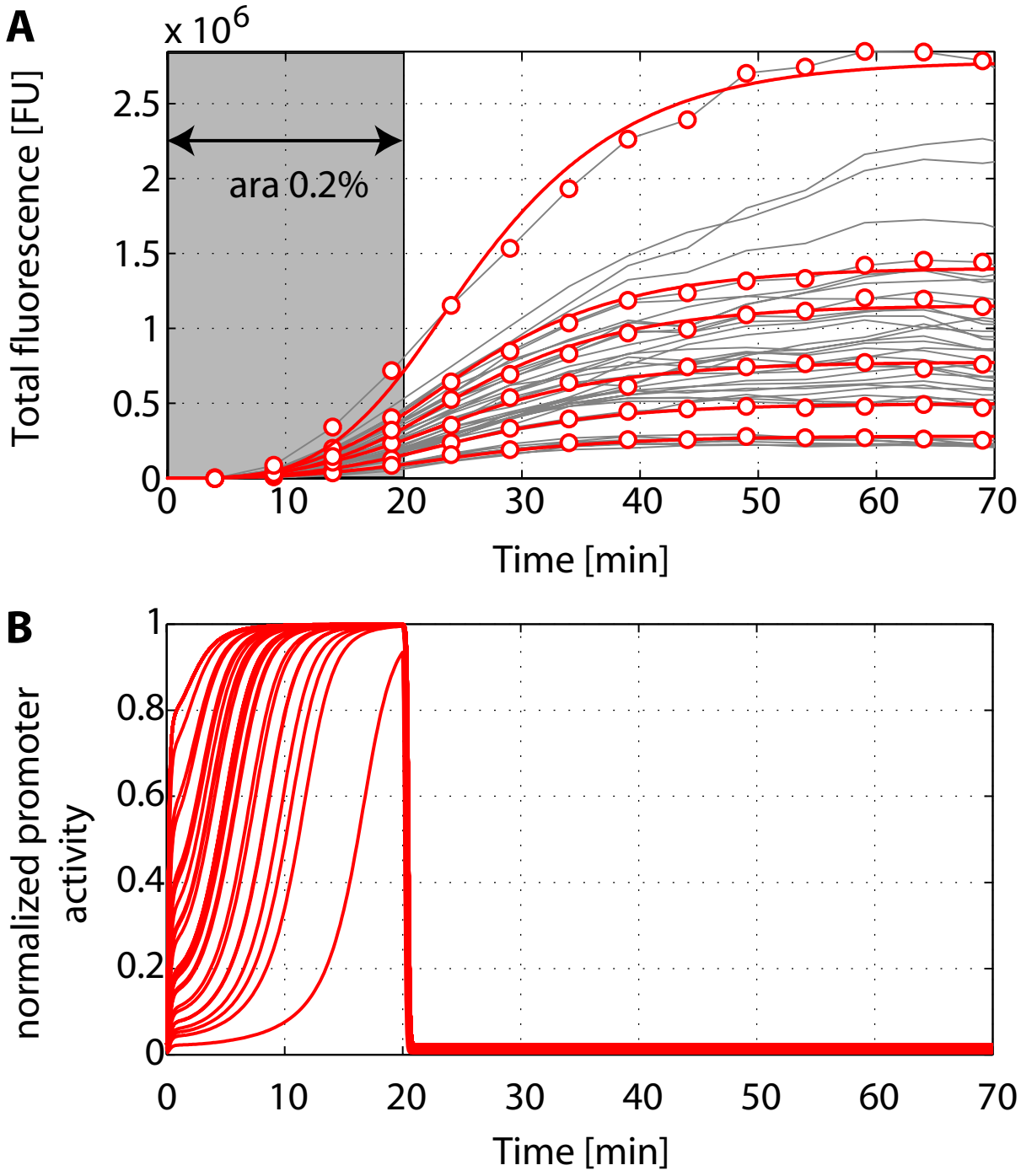

A
x 10^6
Total fluorescence [FU]
2.5
2
1.5
1
0.5
0
ara 0.2%
0
10
20
30
40
50
60
70
Time [min]
B
normalized promoter activity
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
70
Time [min]

Figure3

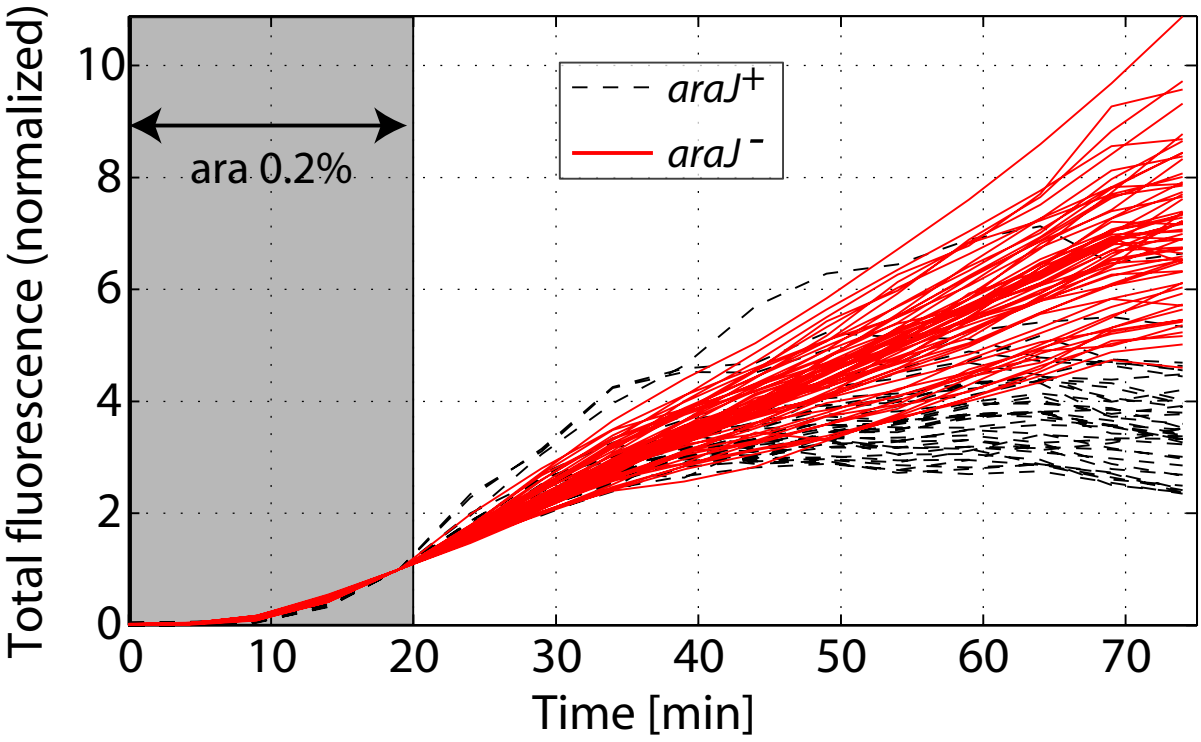

Total fluorescence (normalized)
Time [min]
ara 0.2%
araJ+
araJ-
0
2
4
6
8
10
0
10
20
30
40
50
60
70

Figure4

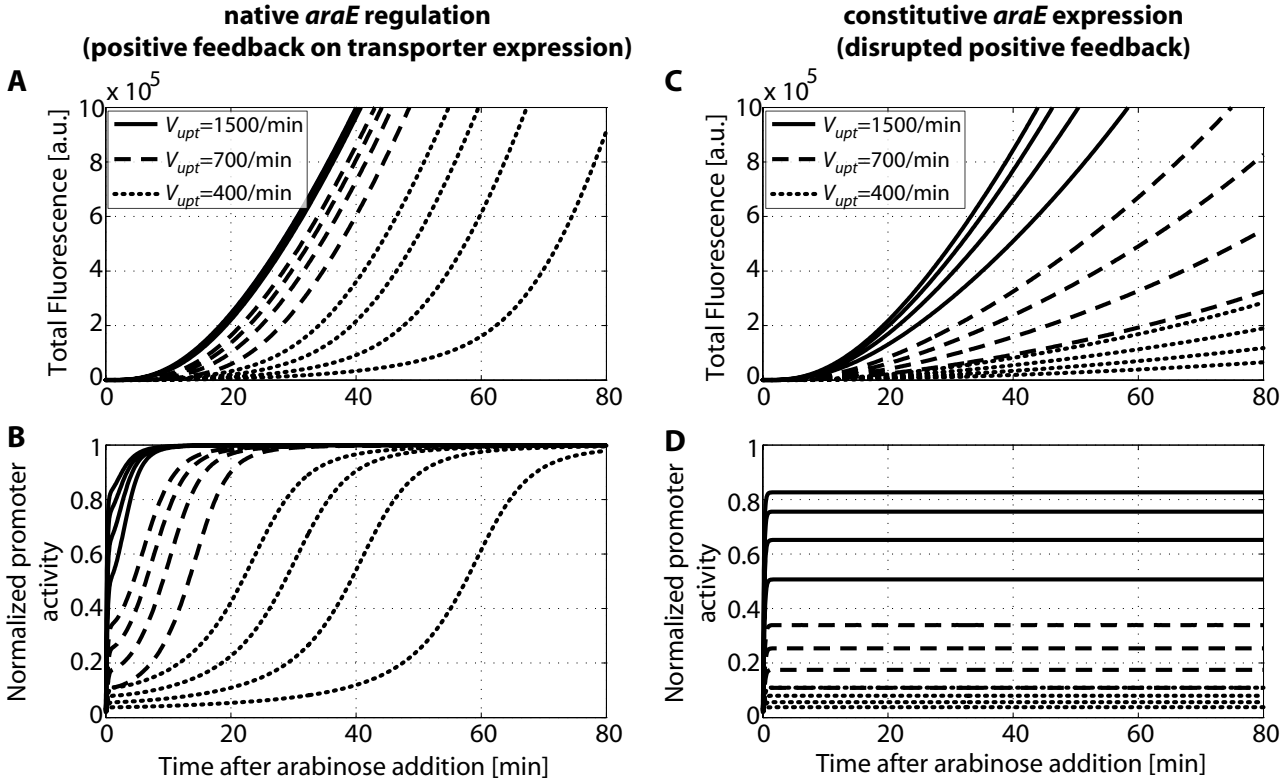
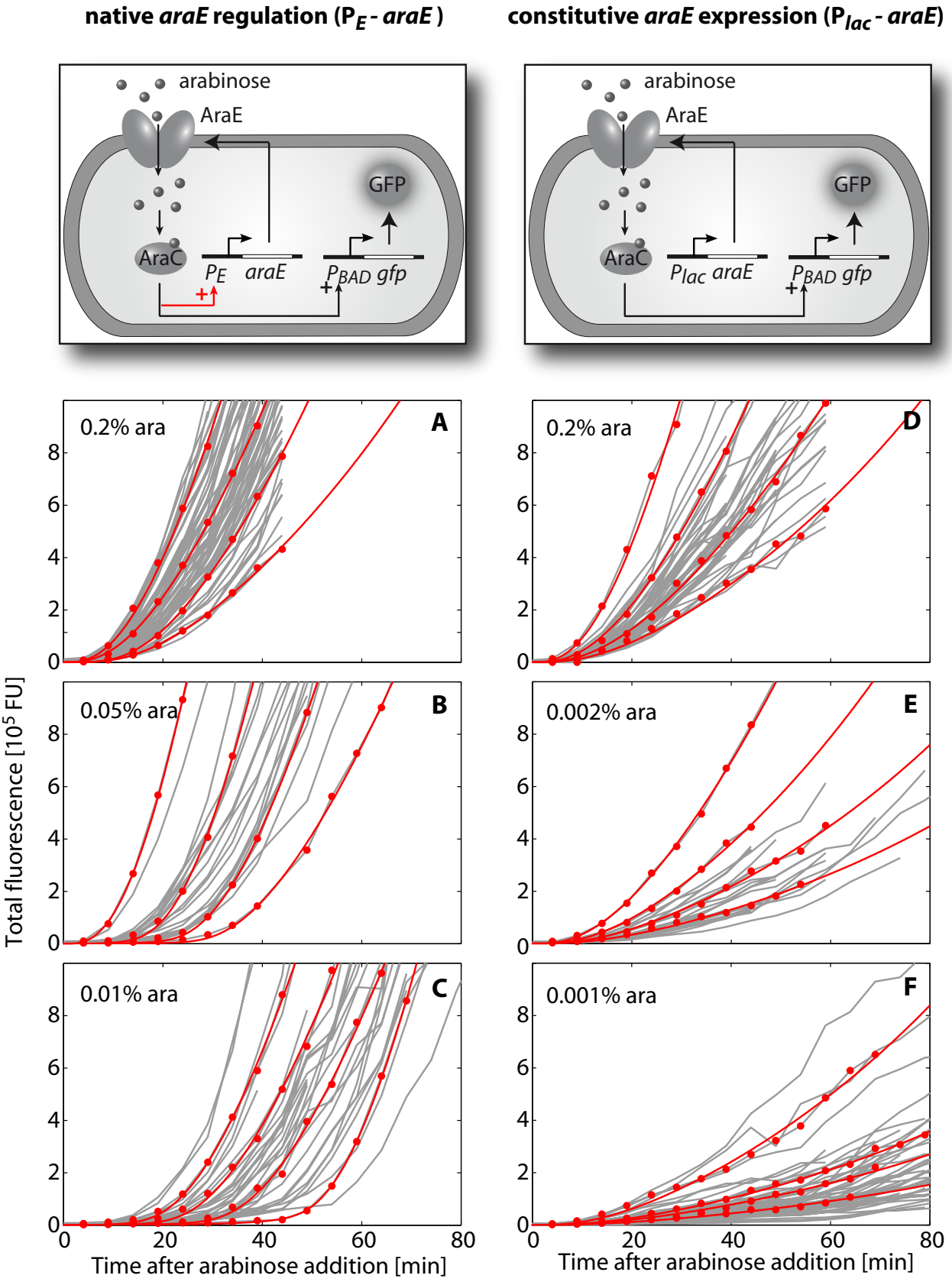

native *araE* regulation
(positive feedback on transporter expression)
constitutive *araE* expression
(disrupted positive feedback)
A
B
C
D
x 10^5
Total Fluorescence [a.u.]
Normalized promoter activity
Time after arabinose addition [min]
V_{upt}=1500/min
V_{upt}=700/min
V_{upt}=400/min

Figure5


native araE regulation (P_E - araE)
constitutive araE expression (P_lac - araE)
arabinose
AraE
GFP
AraC
P_E araE
P_BAD gfp
P_lac araE
0.2% ara
A
0.05% ara
B
0.01% ara
C
D
0.002% ara
E
0.001% ara
F
Total fluorescence [10^5 FU]
Time after arabinose addition [min]

Supplementary Material to:

Quantitative Characterization of Single Cell Switching Dynamics in the Arabinose Utilization System

Judith A. Megerle,† Georg Fritz,‡ Sonja A. Westermayer,† Delia Brick,† Ralf Heermann,§ Kirsten Jung,§ Joachim O. Rädler,† and Ulrich Gerland‡

Judith A. Megerle and Georg Fritz contributed equally to this work.

† Faculty of Physics and CeNS, Ludwig-Maximilians-Universität, Munich, Germany

‡ Arnold Sommerfeld Center for Theoretical Physics and CeNS, Ludwig- Maximilians-Universität, Munich, Germany

§ Center for Integrated Protein Science (CiPSM) at the Department of Biology, Microbiology, Ludwig-Maximilians-Universität, Martinsried, Germany

BACTERIAL STRAINS

Table S1: Bacterial strains used in this study

Strain and description	Relevant genotype	Source or reference
JW1889-1 (Parent of JW1889-3 and JW1889-5)	F$^-$, Δ(*araD-araB*)567, Δ*lacZ*4787(*::rrnB*-3), LAM, Δ*araF*751::*kan*, rph-1, Δ(*rhaD-rhaB*)568, *hsdR*514	Keio collection (Baba 2006) obtained from *`E.coli* genetic stock center' (Yale)
JW1889-3 (native AraE regulation)	F$^-$, *Δ(araD-araB)*567, *ΔlacZ4787*(::*rrnB-3*), LAM, *ΔaraF751::FRT*, *rph-1*, *Δ(rhaD-rhaB)*568, *hsdR514 rpsL150*)	this study
JW1889-5 (*araE* under control of P_{lac})	F$^-$, Δ(*araD-araB*)567, Δ*lacZ*4787(::*rrnB-3*), LAM, Δ*araF751*::FRT, P_{araE}*::*P_{lac} *rph-1*, Δ(*rhaD-rhaB*)568, *hsdR514 rpsL150*)	this study
BW25113 (native *araJ* gene)	F$^-$, *lacI*^{q} *rrnB*T14 Δ*lacZ*WJ16 *hsdR*514 Δ*araBADAH*33 Δ*rhaBADLD*78	(Datsenko 2000) obtained from *`E.coli* genetic stock center' (Yale)
JW0386-1 (Δ*araJ)*	F$^-$, Δ(*araD-araB*)567, Δ*lacZ*4787(*::rrnB*-3), Δ*araJ*760*::kan*, λ$^-$, *rph*-1, Δ(*rhaD-rhaB*)568, *hsdR*514	Keio collection (Baba 2006) obtained from *`E.coli* genetic stock center' (Yale)

CONSTRUCTION OF *E. COLI* STRAINS

The 315 bp intergenic region upstream of *araE* was replaced by the 123 bp intergenic region upstream of *lacZ* in *E. coli* strain JW1889-1 using Red/ET-based *rpsL* genome counter selection (Heermann et al 2008). As a first step, the kanamycine resistance cassette had to be excised from the JW1889-1 genome. Therefore, *E. coli* JW1889-1 was transformed with plasmid pCP20. One colony was picked and incubated in 1 ml LB broth for 3 h at 30°C. Then, temperature was shifted to 37°C and cells were incubated over night. Cells were then plated on LB agar, and single colonies were tested for kanamycine (loss of cassette) and carbenicillin (loss of plasmid pCP20) sensitivity by patching single clones onto LB agar supplemented with kanamycine or carbenicillin, respectively, and the correct clone was named *E. coli* JW1889-2. Then, the *rpsL150* allele was brought into the JW1889-2 genome. *E. coli* strain JW1889-2 was transformed with plasmid pRed/ET(amp) (GeneBridges, Heidelberg). *rpsL150* allele was amplified by PCR with primers homologous to the *rpsL* gene (*rpsL* sense, *rpsL* antisense) using genomic DNA of *E. coli* MC4100 (Casabadan 1976) as template. The linear DNA template was then brought into *E. coli* JW1889-2 by electroporation using a protocol described elsewhere (Heermann et al 2008). Single clones were tested for streptomycine resistance and carbenicillin sensitivity [loss of pRED/ET(amp)] by streaking on LB agar supplemented with the respective antibiotics. The correct clone was named *E. coli* JW1886-3. Then, the intergenic region between *araE* and *kduD* containing the *araE* promoter

region was replaced with the *rpsL*-neo/kan cassette (GeneBridges, Heidelberg) in the JW1886-3 genome. Therefore, the *rpsL*-neo/kan cassette was amplified by PCR using primers with homology arms that are containing the 50 nucleotides upstream and downstream of the *araE/kduD* intergenic target region, respectively (*araE-rpsL*-neo/kan sense, *araE-rpsL*-neo/kan antisense). *E. coli* JW1889-3 was transformed with plasmid pRed/ET(amp). The *araE-rpsL*-neo/kan cassette was then transferred into the *E. coli* JW1889-3 genome by Red/ET recombination. Single clones were tested for kanamycine resistance, and carbenicillin and streptomycine sensitivity. As wild-type *rpsL* transferred with the cassette is dominant over the *rpsL150* allele, cells that have inserted the *rpsL*-neo/kan cassette into the genome become streptomycine sensitive again. The correct strain was designated as *E. coli* JW1889-4. Finally, the *rpsL*-neo/kan cassette was replaced with the *lac* promoter region (intergenic region between *lacI* and *lacY*) using Red/ET based counter selection. Therefore, the *lacI/lacY* intergenic region was amplified by PCR using primers with homology arms containing the 50 nucleotides upstream and downstream of the *araE/kduD* intergenic region (*araE-lacP* sense and *araE-lacP* antisense), and then recombined into the *E. coli* JW1886-4 genome. When the *rpsL*-neo/kan cassette is replaced with the *lacI/lacY* intergenic region, due to the loss of the wild-type *rpsL* gene cells become streptomycine resistant again. Single clones were tested for a kanamycine sensitive, carbenicillin sensitive and streptomycine resistant phenotype. The correct strain was designated as *E. coli* JW1886-5. The correct recombination events were verified at each step by PCR using primers annealing outside of the respective recombined DNA fragment, and by DNA sequencing.

DISCUSSION OF MODELING RATIONALE AND PARAMETER VALUES

For the comparison of our extended model [Eqs. (1)-(7) in the main text] with single-cell fluorescence trajectories, we follow a similar rationale as in ref. (Megerle at al., 2008), where we subsumed all sources of noise during the induction process into the cell-to-cell variability of two key parameters, the delay time τ_d and the protein expression rate α_p. This simplification seemed reasonable, given the significantly lower cell-to-cell variability in all other reaction rates (Megerle at al., 2008). Here we adapt this simplification to our extended model of the induction of the *ara* system and use the basal expression rate of *araE,* $v_{0,e}$, as a proxy for the (inverse) delay time and choose the maximal transcription rate of P_{BAD}, $v_{max,g}$, as a proxy for the protein expression rate. The fixation of both parameters to a static value provides a practicable way to capture stochasticity across a population with a simple deterministic model as long as the observed timespan does not exceed the autocorrelation time (~ cell doubling time) in protein levels significantly, it provides a practicable way to capture stochasticity across a population with a simple deterministic model.

Another key assumption in our analysis is that the arabinose uptake velocity, V_{upt}, scales with the external arabinose concentration, a_{ex}, according to a Michaelis-Menten process, that is, $V_{upt} = V_{max} \times a_{ex}/(K_m+a_{ex})$, where K_m is the Michaelis-Menten constant. Previously, our analysis with the simple gene expression function, in which the transcription rate suddenly switches from `off' to `on' once the internal arabinose threshold is reached, pointed to a K_m of 2.8 mM. This value is about one order of magnitude larger than literature values (Daruwalla et al, 1981). While there might be biological explanations for such a shift under anaerobic conditions (see Megerle et al, 2008), it is also conceivable that it is a mere artifact of our data analysis with the simplified gene expression function. Therefore, here we vary K_m within a physiological range and test which value provides the most coherent explanation for our data. More specifically, we expect that the Michaelis-Menten scaling with external arabinose simultaneously accounts for the modulation in timing (reference strain) and in expression rate (mutant strain), while the distributions of $v_{0,e}$ and $v_{max,g}$ should be independent of the arabinose concentration. Ultimately, the final value of K_m = 300 μM found here is well in the range of previous in vivo determinations (Daruwalla et al, 1981).

The value for the maximal arabinose uptake velocity, V_{max}, was estimated to be in the range of V_{max} = 200-2000 molecules/protein/min (Megerle et al., 2008). While our previous analysis without arabinose efflux required a low number (120 molecules/protein/min) to explain the experimentally observed delay times for reasonable numbers of basally expressed transporters, the presence of the strong efflux in the current model required V_{max} = 2000 molecules/protein/min. This value is also in good agreement with literature values on the lactose/H^+ symporter LacY (1300-3000 molecules/protein/min) (Wright and Overath, 1984; Dornmair et al., 1989).

Reference strain with positive feedback regulation (strain JW1889-3)

The histograms of the estimated parameters for the reference strain are shown in Fig. S1. The histograms of the maximal *gfp* expression rate (Fig. S1D-E) all display comparable average values and standard deviations, as expected. The histograms of the basal *araE* expression rate at lowest (0.01% ara; Fig. S1B) and intermediate arabinose level (0.05% ara; Fig. S1C) also display similar characteristics, but the histogram at highest arabinose level (0.2% ara; Fig. S1A) differs markedly from the others. Here, the basal *araE* expression rate appears bimodal, with an additional peak at the upper bound for the expression rate allowed in our fitting routine (10 transcripts/min). However, at highest external arabinose concentrations the average experimental delay times are very small (cf. Fig. 2A and Table 1 the in main text), such that the relative fluctuations - and thus also the estimation error - are expected to become large. For all arabinose concentrations, the cross-correlations between the two fit parameters (Fig. S2) show that there is little co-variation in both parameters, indicating that both fit parameters can be extracted independently from our data.

Mutant strain without positive feedback regulation (strain JW1889-5)

The histograms of the estimated parameters for the mutant strain with disrupted positive feedback are shown in Fig. S3. Here two clear trends are visible with decreasing arabinose concentration: First, the basal expression rate of *araE* increases on average (Fig. S3A-D) and second, the maximal *gfp* expression rate decreases on average (Fig. S3E-H). However, from our correlation analysis in Fig. S4 we find that there is a strong anticorrelation between both parameters, indicating that only their product is well constrained by our data. This is not surprising, since in the model for the mutant strain both parameters are proportional to the slope of the fluorescence trajectories. Indeed, the histograms of the product of both parameters (Fig. S5) does not display a systematic variation with the external arabinose concentration. However, as in the reference strain, we find that there exists a subpopulation with an overly high product of both parameters, the origin of which is presumably also related to the small delay times at all arabinose concentrations in the mutant strain (cf. Fig. 2D-F and Table 1 the in main text).

SUPPORTING REFERENCES

1. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2:0008.

2. Datsenko, K.A. and Wanner, B, L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. USA 97:6640-6645.

3. Heermann, R., Zeppenfeld, T., Jung, K. (2008). Simple generation of site-directed point mutations in the Escherichia coli chromosome using Red(R)/ET(R) Recombination. Microb. Cell Fact. 7:14.

4. Casadaban, M. J. (1976). Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. J. Mol. Biol. 104:541-555.

5. Megerle, J. A., Fritz, G., Gerland, U., Jung, K. & Rädler, J. O. (2008). Timing and dynamics of single cell gene expression in the arabinose utilization system. Biophys. J. 95:2103-2115.

6. Daruwalla, K. R., Paxton, A. T. & Henderson, P. J. (1981). Energization of the transport systems for arabinose and comparison with galactose transport in *Escherichia coli*. Biochem. J. 200, 611-627.

7. Wright, J. K., and Overath, P. (1984). Purification of the lactose: H+ carrier of *Escherichia coli* and characterization of galactoside binding and transport. Eur. J. Biochem. 138:497-508.

8. Dornmair, K., Overath, P. and Jä hnig, F. (1989). Fast measurement of galactoside transport by lactose permease. J. Biol. Chem. 264:342-346.

Supplementary Figures

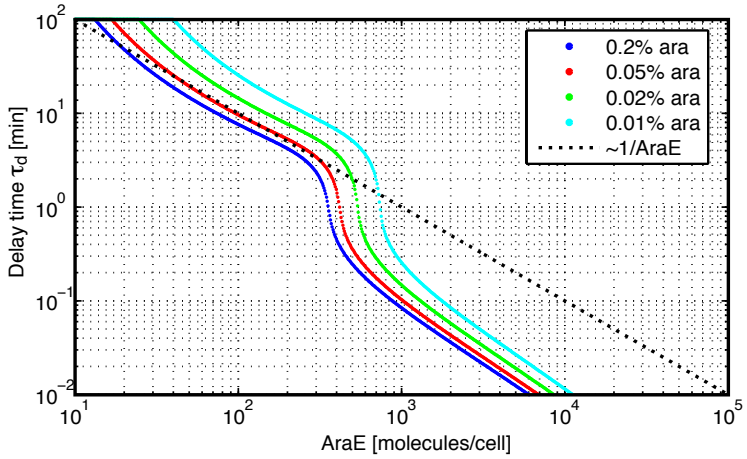


FIGURE S1 Theoretical delay time as a function of the initial number of transporter proteins AraE. The delay time was defined as the time until the P_{BAD} promoter reached 95% of its maximal activity and was estimated from numerical simulations of our model [Eqs. (1)-(7) in the main text] under the initial addition of indicated levels of external arabinose. For all arabinose concentrations we observe a monotonic decrease of the delay time with increasing transporter levels, but with a characteristic kink at around 500 AraE molecules/cell. This strong decrease of the delay time occurs in a regime, in which the high initial number of AraE molecules results in an internal arabinose (quasi-) steady-state level that is already high enough to activate the promoter. For AraE numbers below this kink, the delay time scales to a first approximation inversely with the number of AraE molecules (*dotted line*). The detailed shape of the curves, however, depends on the nature of the nonlinear feedback of AraE on its own synthesis, and we find that the delay time increases stronger than the 1/N-scaling at low AraE numbers.

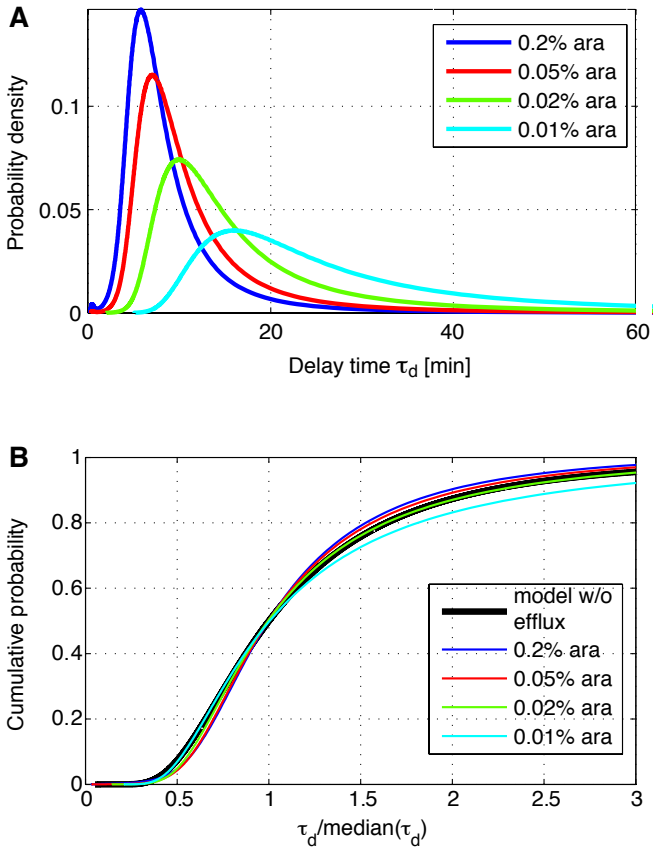


FIGURE S2 Theoretical delay time distributions (**A**) and cumulative probabilities (**B)** at different external arabinose concentrations in our model with arabinose efflux. These data were obtained by using the relation in Fig. S1 to transform a negative binomial distribution of arabinose transporters at the time of sugar addition, *P(n)*, into its corresponding delay time distribution $Q(\tau_d)$, analogous to (Megerle et al., 2008). To that end, we numerically applied the transformation rule $Q(\tau_d)=|dn(\tau_d)/d\tau_d|P(n)$ and used $\langle n\rangle=110$ and $\sigma_n=60$ as parameters for *P(n)*. Due to the inverse scaling of the delay time with the number of AraE proteins on the support of *P(n)* (Fig. S1), the delay time distributions in obtained here (**B**; *colored lines*) only differ within experimental error from our previous results for the delay time distribution in a model without arabinose efflux (**B**; *black solid line*).

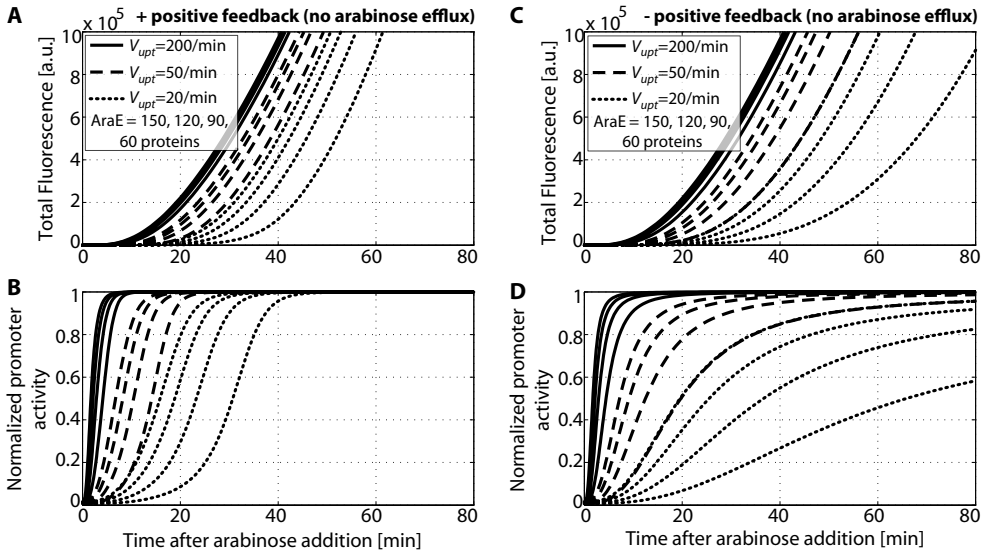


FIGURE S3 Theoretical expression kinetics in a model without arabinose efflux. (**A,B**) The model with positive feedback regulation of arabinose transporters displays qualitatively similar induction kinetics as the model with arabinose efflux (Fig. 4 in the main text): At high V_{upt} trajectories display a homogeneous and rapid response, whereas at low V_{upt} both the average delay as well as the cell-to-cell variability increases. Note that, in order to obtain physiologically relevant delay times, the effective arabinose uptake velocity is about 10 fold lower in the model without arabinose efflux compared to the model with efflux. (**C,D**) Constitutive transporter expression leads to a similar induction kinetics as in (**A,B**). This is due to the fact, that in both models (+/- positive feedback) arabinose uptake is the slow, rate limiting step for gene induction: In each model alike, the rate of arabinose accumulation depends on the initial number of transporters, and consequently, the time for activation of the P_{BAD} promoter is very similar in both cases.

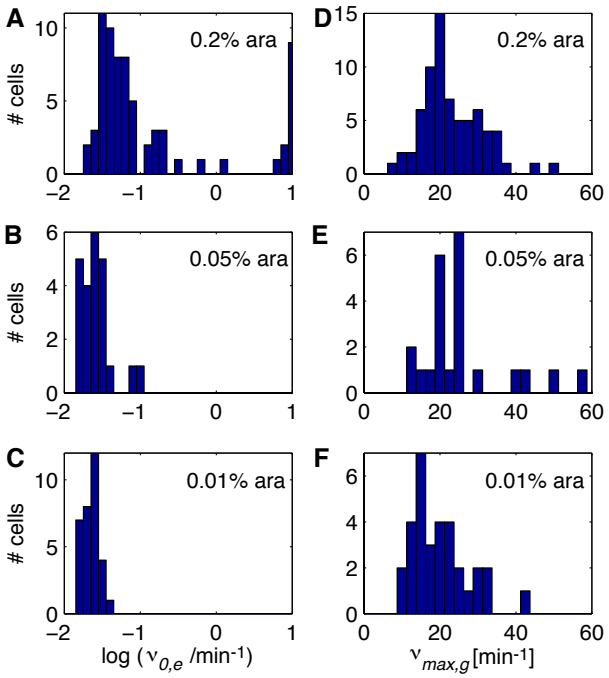


FIGURE S4 Histograms of estimated parameters in the reference strain JW1889-3. (**A**)-(**C**) show the histograms of basal *araE* expression rates and (**D**)-(**F**) the histograms of maximal *gfp* expression rates at indicated arabinose concentrations.

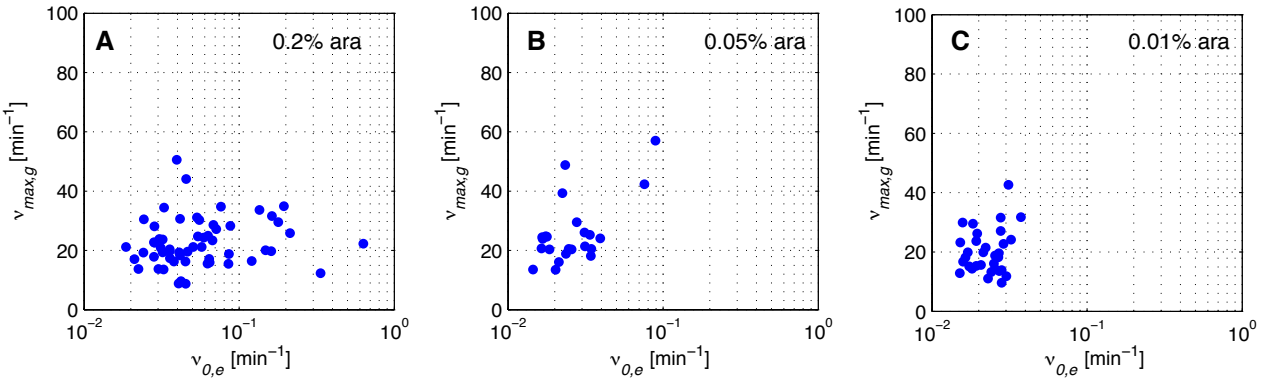


FIGURE S5 Correlations of estimated parameters in the reference strain JW1889-3 at indicated arabinose concentrations.

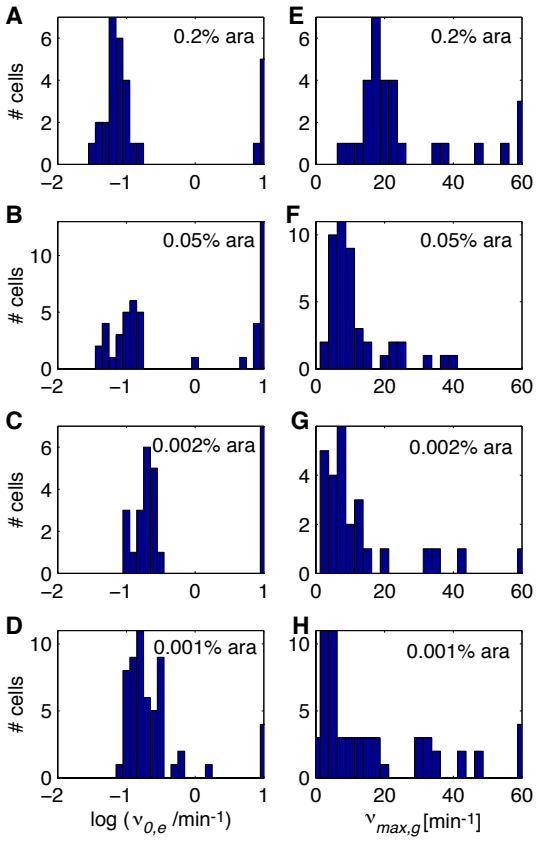


FIGURE S6 Histograms of estimated parameters in strain JW1889-5 with constitutive *araE* expression. (**A**)-(**D**) show the histograms of basal *araE* expression rates and (**E**)-(**H**) the histograms of maximal *gfp* expression rates at indicated arabinose concentrations.

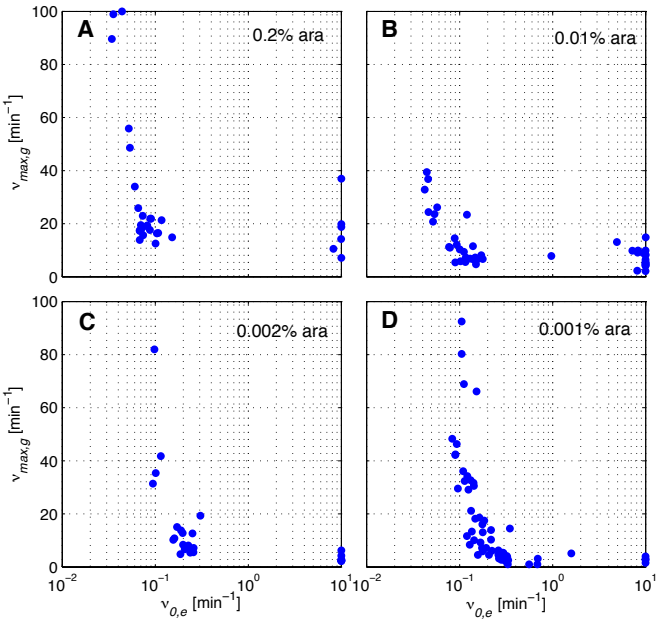


FIGURE S7 Correlations of estimated parameters in strain JW1889-5 with constitutive *araE* expression at indicated arabinose concentrations.

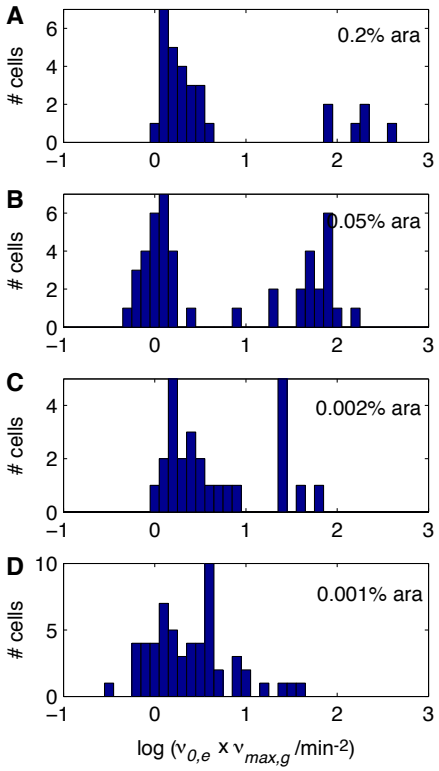


FIGURE S8 Histograms of the product of basal *araE* expression rate and maximal *gfp* expression rate in strain JW1889-5 with constitutive *araE* expression.

B. Full Model for Genetic Latches

To exemplify our modeling approach, we give a list of all reactions contained in our model for the genetic J-K latch in the following. The reaction schemes for the other, simpler latches can be derived from the scheme of the J-K latch by leaving aside individual genes or interactions, while the reaction scheme of the Master-Slave latch is similarly obtained from adding feedback regulation into the regulatory front end, as depicted in Fig. 3.6E. The pictorial reaction scheme for the genetic J-K latch is shown in Fig. B.1 and a comprehensive list of all reactions is given in Table B.1. Here, proteins and their dimers are denoted by capital letters and transcripts of a gene X ($X \in \{A, B, J, K\}$) are denoted by m_X. A gene X is represented by its promoter P_X, which can be occupied by transcriptions factors. Each occupation state of a promoter is represented by an own chemical species for which an empty operator is indicated by $\cdot$ and an occupied operator by the name of the respective transcription factor. In this notation the different binding sites are separated by the symbol $|$. To reduce the number of occupation state combinations, the operator complex for the overlapping heterodimers is separated from the other promoter states and denoted by O. To make transcription nevertheless conditional on a heterodimer operator, we include the respective species for an empty binding site as reactant and product into the corresponding transcription reaction. Finally, the list of parameters used for our simulations is provided in Table B.2.

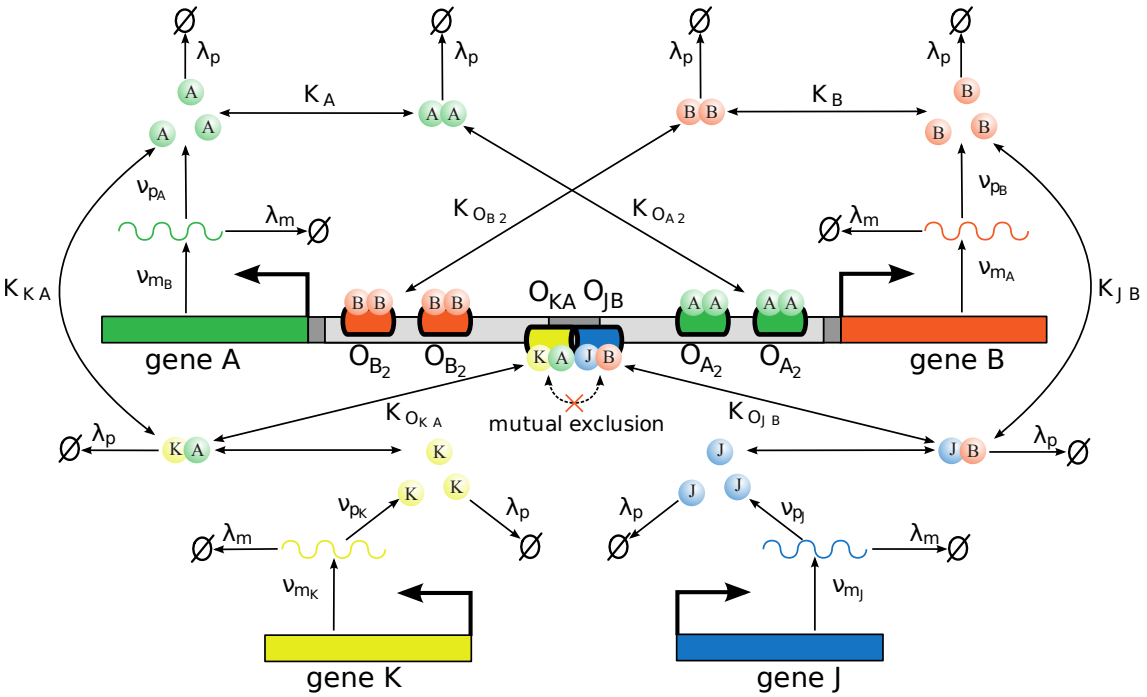


Figure B.1.: Schematic of the reaction network of the JK latch with overlapping heterodimer operators. All reactions contained in the model are depicted together with their rates. On- and off-rates (k^+ and k^-) are combined to dissociation constants $K = k^-/k^+$.

Table B.1.: J-K latch with mutually exclusive binding to O_{KA} and O_{JB}.

Promoter and operator states	
$P_A\|\cdot\|\cdot + B_2 \xrightarrow{k^+_{O_{B2}}} P_A\|B_2\|\cdot$	$P_B\|\cdot\|A_2 \xrightarrow{k^-_{O_{J2}}} P_B + A_2$
$P_A\|B_2\|\cdot \xrightarrow{k^-_{O_{B2}}} P_A\|\cdot\|\cdot + B_2$	$P_B\|\cdot\|A_2 + A_2 \xrightarrow{k^+_{O_{J2}}} P_B\|A_2\|A_2$
$P_A\|\cdot\|\cdot + B_2 \xrightarrow{k^+_{O_{B2}}} P_A\|\cdot\|B_2$	$P_B\|A_2\|A_2 \xrightarrow{k^-_{O_{J2}}} P_B\|\cdot\|A_2 + A_2$
$P_A\|\cdot\|B_2 \xrightarrow{k^-_{O_{B2}}} P_A\|\cdot\|\cdot + B_2$	$P_B\|A_2\|\cdot + A_2 \xrightarrow{k^+_{O_{J2}}} P_B\|A_2\|A_2$
$P_A\|\cdot\|B_2 + B_2 \xrightarrow{k^+_{O_{B2}}} P_A\|B_2\|B_2$	$P_B\|A_2\|A_2 \xrightarrow{k^-_{O_{J2}}} P_B\|A_2\|\cdot + A_2$
$P_A\|B_2\|B_2 \xrightarrow{k^-_{O_{B2}}} P_A\|\cdot\|B_2 + B_2$	$O\|\cdot\|\cdot + KA \xrightarrow{k^+_{O_{KA}}} O\|KA\|\cdot$
$P_A\|B_2\|\cdot + B_2 \xrightarrow{k^+_{O_{B2}}} P_A\|B_2\|B_2$	$O\|KA\|\cdot \xrightarrow{k^-_{O_{KA}}} O\|\cdot\|\cdot + KA$
$P_A\|B_2\|B_2 \xrightarrow{k^-_{O_{B2}}} P_A\|B_2\|\cdot + B_2$	$O\|\cdot\|\cdot + JB \xrightarrow{k^+_{O_{KA}}} O\|\cdot\|JB$
$P_B + A_2 \xrightarrow{k^+_{O_{J2}}} P_B\|A_2\|\cdot$	$O\|\cdot\|JB \xrightarrow{k^-_{O_{KA}}} O\|\cdot\|\cdot + JB$
$P_B\|A_2\|\cdot \xrightarrow{k^-_{O_{J2}}} P_B + A_2$	$O\|KA\|\cdot + KA \xrightarrow{k^+_{KA2}} O\|KA_2\|\cdot$
$P_B + A_2 \xrightarrow{k^+_{O_{J2}}} P_B\|\cdot\|A_2$	$O\|\cdot\|JB + JB \xrightarrow{k^+_{JB2}} O\|\cdot\|JB_2$
Dimerization	
$2A \xrightarrow{k^+_A} A_2$	$J + A \xrightarrow{k^+_J} KA$
$2B \xrightarrow{k^+_B} B_2$	$K + B \xrightarrow{k^+_K} JB$
$A_2 \xrightarrow{k^-_A} 2A$	$KA \xrightarrow{k^-_J} J + A$
$B_2 \xrightarrow{k^-_B} 2B$	$JB \xrightarrow{k^-_K} K + B$
Turnover	
$m_A \xrightarrow{\lambda_{m_A}} \emptyset$	$A_2 \xrightarrow{\lambda_p} \emptyset$
$m_B \xrightarrow{\lambda_{m_B}} \emptyset$	$B_2 \xrightarrow{\lambda_p} \emptyset$
$m_J \xrightarrow{\lambda_m} \emptyset$	$J \xrightarrow{\lambda_p} \emptyset$
$m_K \xrightarrow{\lambda_m} \emptyset$	$K \xrightarrow{\lambda_p} \emptyset$
$A \xrightarrow{\lambda_p} \emptyset$	$KA \xrightarrow{\lambda_p} \emptyset$
$B \xrightarrow{\lambda_p} \emptyset$	$JB \xrightarrow{\lambda_p} \emptyset$
Transcription and translation	
$P_J \xrightarrow{\nu_{m_J}} P_J + m_J$	$m_A \xrightarrow{\nu_{p_A}} m_A + A$
$P_K \xrightarrow{\nu_{m_K}} P_K + m_K$	$m_B \xrightarrow{\nu_{p_B}} m_B + B$
$P_A\|\cdot\|\cdot + O\|\cdot\|\cdot \xrightarrow{\nu_{m_A}} P_A\|\cdot\|\cdot + O\|\cdot\|\cdot + m_A$	$m_J \xrightarrow{\nu_{p_J}} m_J + J$
$P_B\|\cdot\|\cdot + O\|\cdot\|\cdot \xrightarrow{\nu_{m_B}} P_B\|\cdot\|\cdot + O\|\cdot\|\cdot + m_B$	$m_K \xrightarrow{\nu_{p_K}} m_K + K$
$P_A\|\cdot\|\cdot + O\|\cdot\|JB \xrightarrow{\nu_{m_A}} P_A\|\cdot\|\cdot + O\|\cdot\|JB + m_A$	
$P_B\|\cdot\|\cdot + O\|KA\|\cdot \xrightarrow{\nu_{m_B}} P_B\|\cdot\|\cdot + O\|KA\|\cdot + m_B$	

Table B.2.: Parameter values of the J-K latch.

Parameter	Value	Description and References
Transcription		
ν_{m_A}, ν_{m_B}	5 min^{-1}	strong promoter, sequence close to optimum [115]
ν_{m_J}, ν_{m_K}	$(0.01 - 1)$ min^{-1}	inducible promoter [115];
Translation		
ν_{p_A}, ν_{p_B}, ν_{p_J}, ν_{p_K}	2.3 min^{-1}	corresponds to a burst factor of $b = 10$ [178]
Degradation		
λ_m	0.23 min^{-1}	active degradation by RNase [14]; half-life of 3 min
λ_p	0.138 min^{-1}	active proteolysis by SsrA-tags [67]; half life of 5 min
Dimerization		
k_A^+, k_B^+, k_{KA}^+, k_{JB}^+	0.2 $\text{nM}^{-1}\text{min}^{-1}$	assumed to be diffusion limited [22]
k_A^-, k_B^-, k_J^-, k_K^-	2 min^{-1}	typical value for dimerization
Protein-DNA binding		
$k_{O_{A2}}^+$, $k_{O_{B2}}^+$	0.2 $\text{nM}^{-1}\text{min}^{-1}$	assumed to be diffusion limited [22]
$k_{O_{A2}}^-$, $k_{O_{B2}}^-$	2 min^{-1}	modified operator sequence assumed; parameter in the range of *tet*-O_1 binding strength [57]
$k_{O_{KA}}^+$, $k_{O_{JB}}^+$	0.0126 $\text{nM}^{-1}\text{min}^{-1}$	very slow dynamics assumed for on- and off-rate[44, 185]
$k_{O_{KA}}^-$, $k_{O_{JB}}^-$	0.0126 min^{-1}	

Bibliography

[1] H. Abeliovich. An empirical extremum principle for the Hill coefficient in ligand-protein interactions showing negative cooperativity. *Biophys. J.*, 89:76–79, 2005. 17

[2] M. Acar, A. Becskei, and A. van Oudenaarden. Enhancement of cellular memory by reducing stochastic transitions. *Nature*, 435:228–232, 2005. 7, 21, 24, 37, 52

[3] M. Acar, J. T. Mettetal, and A. van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.*, 40:471–475, 2008. 7

[4] S. Ahn, D. D. Ginty, and D. J. Linden. A late phase of cerebellar long-term depression requires activation of CaMKIV and CREB. *Neuron*, 23:559–68, 1999. 24

[5] C. M. Ajo-Franklin, D. A. Drubin, J. A. Eskin, E. P. S. Gee, D. Landgraf, I. Phillips, and P. A. Silver. Rational design of memory in eukaryotic cells. *Genes Dev.*, 21:2271–6, 2007. 21

[6] C. Anetzberger, T. Pirch, and K. Jung. Heterogeneity in quorum sensing-regulated bioluminescence of *Vibrio harveyi*. *Mol. Microbiol.*, 73:267–277, 2009. 7

[7] D. N. Arnosti and M. M. Kulkarni. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell. Biochem.*, 94:890–898, 2005. 36

[8] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305:1622–5, 2004. 7

[9] S. Basu, Y. Gerchman, C. Collins, F. Arnold, and R. Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434:1130–1134, 2005. 30

[10] S. Bearson, B. Bearson, and J. W. Foster. Acid stress responses in enterobacteria. *FEMS Microbiol. Lett.*, 147:173–180, 1997. 11

[11] H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey. Experimental evolution of bet hedging. *Nature*, 462:90–U97, 2009. 7

[12] A. Becskei, B. B. Kaufmann, and A. van Oudenaarden. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.*, 37: 937–944, 2005. 54

[13] O. G. Berg. A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.*, 71:587–603, 1978. 42

[14] J. A. Bernstein, A. B. Khodursky, P. H. Lin, S. Lin-Chao, and S. N. Cohen. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA*, 99:9697–9702, 2002. 152

[15] S. R. Biggar and G. R. Crabtree. Cell signaling can direct either binary or graded transcriptional responses. *EMBO J.*, 20:3167–76, 2001. 7, 52

[16] L. Bintu, N. E. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.*, 15:125–135, 2005. 21, 29

[17] L. Bintu, N. E. Buchler, H. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, 15:116–124, 2005. 21, 29

[18] S. L. Blethen, E. A. Boeker, and E. E. Snell. Argenine decarboxylase from *Escherichia coli*. I. Purification and specificity for substrates and coenzyme. *J. Biol. Chem.*, 243: 1671–1677, 1968. 18

[19] E. A. Boeker and E. E. Snell. *Amino acid decarboxylases*, volume 6, chapter 7, pages 217–253. Academic Press, New York, 3rd edition, 1972. 11

[20] S. Bost, F. Silva, and D. Belin. Transcriptional activation of *ydeA*, which encodes a member of the major facilitator superfamily, interferes with arabinose accumulation and induction of the *Escherichia coli* arabinose P_{BAD} promoter. *J. Bacteriol.*, 181: 2185–2191, 1999. 41

[21] H. Bremer and P. P. Dennis. *Physiology of the bacterial cell*, chapter Modulation of chemical composition and other parameters of the cell by growth rate, page 1553. Sinauer Associates, Inc., Sunderland, Massachusetts, 1990. 1

[22] R. F. Bruinsma. Physics of protein-DNA interaction. *Physica A*, 313:211–237, 2002. 152

[23] N. E. Buchler, U. Gerland, and T. Hwa. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA*, 100:5136–5141, 2003. 5, 21, 22, 23

[24] D. R. Burrill and P. A. Silver. Making cellular memories. *Cell*, 140:13–18, 2010. 27

[25] P. H. Calcott and J. R. Postgate. On substrate-accelerated death in *Klebsiella aerogenes*. *J. Gen. Microbiol.*, 70:115–22, 1972. 53

[26] O. Canela-Xandri, F. Sagués, R. Reigada, and J. Buceta. A spatial toggle switch drives boundary formation in development. *Biophys. J.*, 95:5111–5120, 2008. 21

[27] S. Carole, S. Pichoff, and J.-P. Bouche. *Escherichia coli* gene *ydeA* encodes a major facilitator pump which exports L-arabinose and isopropyl-β-D-thiogalactopyranoside. *J. Bacteriol.*, 181:5123–5125, 1999. 41

[28] J. Casadesus and R. D'Ari. Memory in bacteria and phage. *Bioessays*, 24(6):512–518, 2002. 6

[29] J. Casadesús and D. Low. Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.*, 70:830–56, 2006. 6, 52

[30] M.-P. Castanié-Cornet, K. Cam, B. Bastiat, A. Cros, P. Bordes, and C. Gutierrez. Acid stress response in *Escherichia coli*: mechanism of regulation of *gadA* transcription by RcsB and GadE. *Nucl. Acids Res.*, 2010. 27

[31] D.-E. Chang, S. Leung, M. R. Atkinson, A. Reifler, D. Forger, and A. J. Ninfa. Building biological memory by linking positive feedback loops. *Proc. Natl. Acad. Sci. USA*, 107: 175–80, 2010. 21

[32] A. G. Chapman, L. Fall, and D. E. Atkinson. Adenylate energy charge in *Escherichia coli* during growth and starvation. *J. Bacteriol.*, 108:1072–86, 1971. 53

[33] A. Chastanet, D. Vitkup, G.-C. Yuan, T. M. Norman, J. S. Liu, and R. M. Losick. Broadly heterogeneous activation of the master regulator for sporulation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 107:8486–8491, 2010. 54

[34] J. Cooke and E. C. Zeeman. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *J. Theor. Biol.*, 58:455–76, 1976. 34

[35] R. S. Cox, M. G. Surette, and M. B. Elowitz. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.*, 3:145, 2007. 21

[36] K. R. Daruwalla, A. T. Paxton, and P. J. Henderson. Energization of the transport systems for arabinose and comparison with galactose transport in *Escherichia coli*. *Biochem. J.*, 200:611–627, 1981. 40, 42, 43

[37] E. H. Davidson. Network design principles from the sea urchin embryo. *Curr. Opin. Genet. Dev.*, 19:535–40, 2009. 21

[38] I. G. de Jong, J.-W. Veening, and O. P. Kuipers. Heterochronic phosphorelay gene expression as a source of heterogeneity in *Bacillus subtilis* spore formation. *J. Bacteriol.*, 192:2053–67, 2010. 54

[39] E. Dekel and U. Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436:588–92, 2005. 1, 5, 8, 45, 52

[40] E. Dekel, S. Mangan, and U. Alon. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys. Biol.*, 2:81–88, 2005. 45

[41] C. L. Dell, M. N. Neely, and E. R. Olson. Altered pH and lysine signalling mutants of *cadC*, a gene encoding a membrane-bound transcritpional activator of the *Escherichia coli cadBA* operon. *Mol. Microbiol.*, 14:7–16, 1994. 12

[42] M.-L. Dequéant and O. Pourquié. Segmental patterning of the vertebrate embryonic axis. *Nat. Rev. Genet.*, 9:370–82, 2008. 21, 34, 35

[43] I. B. Dodd, M. A. Micheelsen, K. Sneppen, and G. Thon. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*, 129:813–22, 2007. 25

[44] H. Dong, L. Nilsson, and C. G. Kurland. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.*, 177: 1497–504, 1995. 1, 152

[45] D. Dykhuizen and D. Hartl. Transport by the lactose permease of *Escherichia coli* as the basis of lactose killing. *J. Bacteriol.*, 135:876–82, 1978. 53

[46] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1:54–75, 1986. 16

[47] A. Eichinger, I. Haneburger, C. Koller, K. Jung, and A. Skerra. Crystal structure of the sensory domain of *Escherichia coli* CadC, a member of the ToxR-like protein family. *Prot. Sci.*, 20:656–669, 2011. 9, 17

[48] A. Eldar and M. B. Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467: 167–173, 2010. 7, 37

[49] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000. ISSN 0028-0836. 30

[50] A. Farewell, K. Kvint, and T. Nyström. Negative regulation by RpoS: a case of sigma factor competition. *Mol. Microbiol.*, 29:1039–1051, 1998. 4

[51] J. E. Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.*, 14:140–148, 2002. 5, 27, 37

[52] J. W. Foster. *Microbial responses to acid stress*, chapter 7, pages 99–115. ASM Press, Washington, D.C., 2000. 18

[53] J. W. Foster. *Escherichia coli* acid resistance: Tales of an amateur acidophile. *Nat. Rev. Microb.*, 2:898–907, 2004. ISSN 1740-1526. 10, 11

[54] P. François, V. Hakim, and E. D. Siggia. Deriving structure from evolution: metazoan segmentation. *Mol. Syst. Biol.*, 3:154, 2007. 34

[55] D. Fraser and M. Kærn. A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol. Microbiol.*, 71:1333–1340, 2009. 7

[56] H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen. Noise minimization in eukaryotic gene expression. *PLoS Biology*, 2:e137, 2004. 7

[57] G. Fritz, N. E. Buchler, T. Hwa, and U. Gerland. Designing sequential transcription logic: a simple genetic circuit for conditional memory. *Syst. Synth. Biol.*, 1:89–98, 2007. 21, 27, 32, 152

[58] G. Fritz, C. Koller, K. Burdack, L. Tetsch, I. Haneburger, K. Jung, and U. Gerland. Induction kinetics of a conditional pH stress response system in *Escherichia coli*. *J. Mol. Biol.*, 393:272–286, 2009. 9

[59] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S. Lee, and J. C. Liao. A synthetic gene-metabolic oscillator. *Nature*, 435:118–122, 2005. ISSN 0028-0836. 30

[60] E. F. Gale and H. M. R. Epps. The effect of the pH of the medium during growth on the enzymic activities of bacteria (*Escherichia coli* and *Micrococcus lysodeikticus*) and the biological significance of the changes produced. *Biochem. J.*, 36:600–618, 1942. 10, 11

[61] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, 2000. ISSN 0028-0836. 21, 24, 27, 29, 30

[62] F. Geier, G. Fengos, F. Felizzi, and D. Iber. *Analysing and constraining signaling networks: parameter estimation for the user*. Humana Press, in press. 13, 16

[63] N. Geisel, J. M. G. Vilar, and J. M. Rubi. Optimal resting-growth strategies of microbial populations in fluctuating environments. *PLoS ONE*, 6:e18622, 2011. 1

[64] G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17:883–893, 1980. 14

[65] C. Gomez, E. M. Ozbudak, J. Wunderlich, D. Baumann, J. Lewis, and O. Pourquié. Control of segment number in vertebrate embryos. *Nature*, 454:335–9, 2008. 21, 34

[66] M. Goodson and R. J. Rowbury. Habituation to normally lethal acidity by prior growth of *Escherichia coli* at a sub-lethal acid pH value. *Lett. Appl. Microbiol.*, 8(2):77–79, 1989. 11

[67] S. Gottesman, E. Roche, Y. N. Zhou, and R. T. Sauer. The clpxp and clpap proteases degrade proteins with carboxy-terminal peptide tails added by the ssra-tagging system. *Genes Dev.*, 12:1338–1347, 1998. 152

[68] C. C. Guet, M. B. Elowitz, W. Hsing, and S. Leibler. Combinatorial synthesis of genetic networks. *Science*, 296:1466–70, 2002. 21

[69] J. R. Haanstra, A. van Tuijl, P. Kessler, W. Reijnders, P. A. M. Michels, H. V. Westerhoff, M. Parsons, and B. M. Bakker. Compartmentation prevents a lethal turboexplosion of glycolysis in trypanosomes. *Proc. Natl. Acad. Sci. USA*, 105:17718–23, 2008. 54

[70] T. S. Ham, S. K. Lee, J. D. Keasling, and A. P. Arkin. Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLoS ONE*, 3:e2815, 2008. 21

[71] I. Haneburger, A. Eichinger, A. Skerra, and K. Jung. New Insights into the Signaling Mechanism of the pH-responsive, Membrane-integrated Transcriptional Activator CadC of Escherichia coli. *J. Biol. Chem.*, 286:10681–10689, 2011. 9, 12, 17

[72] I. Haneburger, G. Fritz, N. Jurkschat, L. Tetsch, A. Eichinger, A. Skerra, U. Gerland, and K. Jung. The feedback-inhibitor cadaverine suppresses pH response by binding to the pH susceptible site of CadC in *Escherichia coli. submitted*, 2011. 9

[73] W. Hendrickson, C. Stoner, and R. Schleif. Characterization of the *Escherichia coli araFGH* and *araJ* promoters. *J. Mol. Biol.*, 215:497–510, 1990. 41

[74] L. Herrgen, S. Ares, L. G. Morelli, C. Schröter, F. Jülicher, and A. C. Oates. Intercellular coupling regulates the period of the segmentation clock. *Curr. Biol.*, 20:1244–53, 2010. 34, 35

[75] R. W. Hogg. L-Arabinose transport and the L-arabinose binding protein of *Escherichia coli*. *J. Supramol. Struct.*, 6:411–417, 1977. 40

[76] S. Hohmann. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol. Mol. Biol. Rev.*, 66:300–372, 2002. 5

[77] M. Hollis, D. Valenzuela, D. Pioli, R. Wharton, and M. Ptashne. A repressor heterodimer binds to a chimeric operator. *Proc. Natl. Acad. Sci. USA*, 85:5834–5838, 1988. 27

[78] B. F. Horazdovsky and R. W. Hogg. High-affinity L-arabinose transport operon. Gene product expression and mRNAs. *J. Mol. Biol.*, 197:27–35, 1987. 40

[79] F. J. Isaacs, J. Hasty, C. R. Cantor, and J. Collins. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA*, 100:7714–7719, 2003. 21

[80] R. Iyer, C. Williams, and C. Miller. Arginine-agmatine antiporter in extreme acid resistance in *Escherichia coli*. *J. Bacteriol.*, 185:6556–6561, 2003. 11

[81] K. Jaqaman and G. Danuser. Linking data to models: data regression. *Nat. Rev. Mol. Cell Biol.*, 7:813–819, 2006. 13, 14

[82] T. Kalisky, E. Dekel, and U. Alon. Cost-benefit theory and optimal design of gene regulation functions. *Phys. Biol.*, 4:229–45, 2007. 1, 5, 45, 52

[83] S. Kaplan, A. Bren, A. Zaslaver, E. Dekel, and U. Alon. Diverse two-dimensional input functions control bacterial sugar genes. *Mol. Cell.*, 29:786–92, 2008. 21

[84] R. H. Katz. *Contemporary Logic Design*. Benjamin Cummings, 1994. 22, 27, 32

[85] B. B. Kaufmann, Q. Yang, J. T. Mettetal, and A. van Oudenaarden. Heritable stochastic switching revealed by single-cell genealogy. *PLoS Biol.*, 5:e239, 2007. 6, 7, 52

[86] J. L. Kelly. A new interpretation of information rate. *Bell. Syst. Tech. J.*, 35:917–926, 1956. 52

[87] I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp. EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucl. Acids Res.*, 39:D583–90, 2011. 40

[88] A. Khlebnikov, O. Risa, T. Skaug, T. A. Carrier, and J. D. Keasling. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.*, 182:7029–7034, 2000. 38

[89] A. Khlebnikov, K. A. Datsenko, T. Skaug, B. L. Wanner, and J. D. Keasling. Homogeneous expression of the P-BAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology*, 147:3241–3247, 2001.

[90] A. Khlebnikov, T. Skaug, and J. D. Keasling. Modulation of gene expression from the arabinose-inducible araBAD promoter. *J. Ind. Microbiol. Biotechnol.*, 29:34–37, 2002. 38

[91] H. Kobayashi, M. Kaern, M. Araki, K. Chung, and T. Gardner. Programmable cells: Interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA*, 101: 8414–8419, 2004. 21

[92] A. L. Koch and C. H. Wang. How close to the theoretical diffusion limit do bacterial uptake systems function? *Arch. Microbiol.*, 131:36–42, 1982. 45

[93] K. Kolodrubetz and R. Schleif. Regulation of the L-arabinose transport operons in *Escherichia coli*. *J. Mol. Biol.*, 151:215–227, 1981. 39

[94] E. Korobkova, T. Emonet, J. M. G. Vilar, T. S. Shimizu, and P. Cluzel. From molecular noise to behavioural variability in a single bacterium. *Nature*, 428:574–578, 2004. 7

[95] C. Kreutz and J. Timmer. Systems biology: experimental design. *FEBS J.*, 276:923–942, 2009. 13

[96] T. A. Krulwich, G. Sachs, and E. Padan. Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.*, 9:330–343, 2011. 11

[97] T. Kuhlman, Z. Zhang, M. Saier, and T. Hwa. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 104:6043–8, 2007. 5, 21, 46

[98] C. Küper and K. Jung. CadC-mediated activation of the *cadBA* promoter in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.*, 10:26–39, 2005. 12

[99] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309:2075–2078, 2005. 6, 7, 37

[100] G. I. Lang, A. W. Murray, and D. Botstein. The cost of gene expression underlies a fitness trade-off in yeast. *Proc. Natl. Acad. Sci. USA*, 106(14):5755–60, Apr 2009. 1

[101] M. T. Laub and M. Goulian. Specificity in two-component signal transduction pathways. *Annu. Rev. Genet.*, 41:121–145, 2007. 12

[102] Y. H. Lee, J. H. Kim, I. S. Bang, and Y. K. Park. The membrane-bound transcriptional regulator CadC is activated by proteolytic cleavage in response to acid stress. *J. Bacteriol.*, 190:5120–5126, 2008. 12

[103] S. Legewie, D. Dienst, A. Wilde, H. Herzel, and I. M. Axmann. Small RNAs establish delays and temporal thresholds in gene expression. *Biophys. J.*, 95:3232–3238, 2008. 5

[104] R. E. Lenski, V. Souza, L. P. Duong, Q. G. Phan, T. N. Nguyen, and K. P. Bertrand. Epistatic effects of promoter and repressor functions of the Tn10 tetracycline-resistance operon of the fitness of *Escherichia coli*. *Mol. Ecol.*, 3:127–35, 1994. 46

[105] R. E. Lenski, J. A. Mongold, P. D. Sniegowski, M. Travisano, F. Vasi, P. J. Gerrish, and T. M. Schmidt. Evolution of competitive fitness in experimental populations of *E. coli*: what makes one genotype a better competitor than another? *Antonie van Leeuwenhoek*, 73:35–47, 1998. 1

[106] E. Levine, Z. Zhang, T. Kuhlman, and T. Hwa. Quantitative characteristics of gene regulation by small RNA. *PLoS Biol.*, 5:e229, 2007. 5

[107] J. Lewis. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.*, 13:1398–1408, 2003. 5

[108] J. Lewis, A. Hanisch, and M. Holder. Notch signaling, the segmentation clock, and the patterning of vertebrate somites. *J. Biol.*, 8:44, 2009. 35

[109] D. J. Linden. A protein synthesis-dependent late phase of cerebellar long-term depression. *Neuron*, 17:483–90, 1996. 24

[110] J. E. Lisman. A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proc. Natl. Acad. Sci. USA*, 82:3055–7, 1985. 24

[111] D. López and R. Kolter. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. *FEMS Microbiol. Rev.*, 34:134–49, 2010. 6

[112] D. López, H. Vlamakis, and R. Kolter. Generation of multiple cell types in *Bacillus subtilis*. *FEMS Microbiol. Rev.*, 33:152–163, 2009. 7

[113] R. Losick and C. Desplan. Stochasticity and cell fate. *Science*, 320:65–8, 2008. 7

[114] C. Lou, X. Liu, M. Ni, Y. Huang, Q. Huang, L. Huang, L. Jiang, D. Lu, M. Wang, C. Liu, D. Chen, C. Chen, X. Chen, L. Yang, H. Ma, J. Chen, and Q. Ouyang. Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol. Syst. Biol.*, 6: 350, 2010. 21

[115] R. Lutz and H. Bujard. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I-1-I-2 regulatory elements. *Nucl. Acids Res.*, 25:1203–1210, 1997. 38, 152

[116] P. C. Maloney and B. Rotman. Distribution of suboptimally induced β-D-galactosidase in *Escherichia coli*. The enzyme content of individual cells. *J. Mol. Biol.*, 73:77–91, 1973. 7

[117] L. M. Maurer, E. Yohannes, S. S. Bondurant, M. Radmacher, and J. L. Slonczewski. pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *J. Bacteriol.*, 187:304–319, 2005. 11

[118] H. H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269: 650–6, 1995. 21

[119] J. A. Megerle, G. Fritz, U. Gerland, K. Jung, and J. O. Rädler. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.*, 95:2103–2115, 2008. 21, 37

[120] J. A. Megerle, G. Fritz, S. A. Westermayer, D. Brick, R. Heermann, K. Jung, J. O. Rädler, and U. Gerland. Quantitative characterization of single cell switching dynamics in the arabinose utilization systems. *submitted*, 2011. 37

[121] B. A. Mello and Y. Tu. An allosteric model for heterogeneous receptor complexes: understanding bacterial chemotaxis responses to multiple stimuli. *Proc. Natl. Acad. Sci. USA*, 102:17354–17359, 2005. 5

[122] S. Y. Meng and G. N. Bennett. Nucleotide sequence of the *Escherichia coli cad* operon: a system for neutralization of low extracellular pH. *J. Bacteriol.*, 174:2659–2669, 1992. 12

[123] A. Mitchell, G. H. Romano, B. Groisman, A. Yona, E. Dekel, M. Kupiec, O. Dahan, and Y. Pilpel. Adaptive prediction of environmental changes by microorganisms. *Nature*, 460:220–4, 2009. 6

[124] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.*, 13:2467–2474, 2003. 14

[125] J. Monod. The growth of bacterial cultures. *Annu. Rev. Microbiol.*, 3:371–394, 1949. 1, 4

[126] J. Monod. From enzymatic adaptation to allosteric transitions. *Science*, 154:475–83, 1966. 5, 37

[127] R. M. Morgan-Kiss, C. Wadler, and J. E. J. Cronan. Long-term and homogeneous regulation of the *Escherichia coli araBAD* promoter by use of a lactose transporter of relaxed specificity. *Proc. Natl. Acad. Sci. USA*, 99:7373–7377, 2002. 38

[128] M. N. Neely and E. R. Olson. Kinetics of expression of the *Escherichia coli cad* operon as a function of pH and lysine. *J. Bacteriol.*, 178:5522–5528, 1996. 9, 12

[129] M. N. Neely, C. L. Dell, and E. R. Olson. Roles of LysP and CadC in mediating the lysine requirement for acid induction of the *Escherichia coli cad* operon. *J. Bacteriol.*, 176:3278–3285, 1994. 12

[130] A. Novick and M. Weiner. Enzyme induction as an all-or-none phenomenon. *Proc. Natl. Acad. Sci. USA*, 43:553–566, 1957. 7, 8, 21, 24, 37, 46, 52

[131] C. P. Novotny and E. Englesberg. The L-arabinose permease system in *Escherichia coli* B/r. *Biochim. Biophys. Acta*, 117:217–230, 1966. 40, 43

[132] H. Ogasawara and M. Kawato. Bistable switches for synaptic plasticity. *Sci. Signal*, 2: pe7, 2009. 24

[133] E. Oxman, U. Alon, and E. Dekel. Defined order of evolutionary adaptations: experimental evidence. *Evolution*, 62:1547–54, 2008. 46

[134] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, 427:737–740, 2004. 5, 7, 8, 21, 24, 37, 46, 52

[135] S. S. Pao, I. T. Paulsen, and M. H. Saier. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.*, 62:1–34, 1998. 40

[136] P. Paszek, S. Ryan, L. Ashall, K. Sillitoe, C. V. Harper, D. G. Spiller, D. A. Rand, and M. R. H. White. Population robustness arising from cellular heterogeneity. *Proc. Natl. Acad. Sci. USA*, 107:11644–9, 2010. 7

[137] J. Paulsson and M. Ehrenberg. Noise in a minimal regulatory network: plasmid copy number control. *Q. Rev. Biophys.*, 34:1–59, 2001. 5

[138] I. S. Peter and E. H. Davidson. Modularity and design principles in the sea urchin embryo gene regulatory network. *FEBS Lett.*, 583:3948–58, 2009. 21

[139] J. D. Pfau and R. K. Taylor. Mutations in *toxR* and *toxS* that separate transcriptional activation from DNA binding at the cholera toxin gene promoters. *J. Bacteriol.*, 180: 4724–4733, 1998. 12

[140] R. Phillips and R. Milo. A feeling for the numbers in biology. *Proc. Natl. Acad. Sci. USA*, 106:21465–21471, 2009. 4

[141] F. J. Poelwijk, M. G. J. de Vos, and S. J. Tans. Tradeoffs and optimality in the evolution of gene regulation. *Cell*, 146:462–470, 2011. 1, 5

[142] J. R. Postgate and J. R. Hunter. Acceleration of bacterial death by grown substrates. *Nature*, 198:273, 1963. 53

[143] W. H. Press, S. A. Teukolsky, and W. T. Vetterling. Confidence Limits on Estimated Model Parameters. In *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, 1999. 15

[144] A. V. Probst, E. Dunleavy, and G. Almouzni. Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.*, 10:192–206, 2009. 25

[145] M. Ptashne and A. Gann. *Genes and Signals*. Cold Spring Harbor Laboratory Press, 2002. 37

[146] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135:216–26, 2008. 7, 52

[147] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmuller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25:1923–1929, 2009. 14, 15, 16

[148] T. Reeder and R. Schleif. Mapping, sequence, and apparent lack of function of *araJ*, a gene of the *Escherichia coli* arabinose regulon. *J. Bacteriol.*, 173:7765–7771, 1991. 38, 40

[149] L. Robert, G. Paul, Y. Chen, F. Taddei, D. Baigl, and A. B. Lindner. Pre-dispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch. *Mol. Syst. Biol.*, 6:357, 2010. 6

[150] J. B. Russell and G. M. Cook. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Rev.*, 59:48–62, 1995. 2

[151] D. L. Sabo, E. Boeker, B. Byers, H. Waron, and E. H. Fischer. Purification and physical properties of inducible *Escherichia coli* lysine decarboxylase. *Biochemistry*, 13:662–670, 1974. 11, 12

[152] H. Samartzidou and A. H. Delcour. Excretion of endogenous cadaverine leads to a decrease in porin-mediated outer membrane permeability. *J. Bacteriol.*, 181:791–798, 1999. 11

[153] H. Samartzidou, M. Mehrazin, Z. Xu, M. J. Benedik, and A. H. Delcour. Cadaverine inhibition of porin plays a role in cell survival at acidic pH. *J. Bacteriol.*, 185:13–19, 2003. 11

[154] M. Schaechter, O. Maaloe, and N. O. Kjelgaard. Dependency on medium and temperature of cell size and chemical composition during balanced grown of *Salmonella typhimurium*. *J. Gen. Microbiol.*, 19:592–606, 1958. 1

[155] R. Schleif. Induction of the L-arabinose operon. *J. Mol. Biol.*, 46:197–199, 1969. 46

[156] R. Schleif. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet.*, 16: 559–565, 2000. 38

[157] M. Scott and T. Hwa. Bacterial growth laws and their applications. *Curr. Opin. Biotechnol.*, 22:559–565, 2011. 1, 3

[158] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330:1099–1102, 2010. 1, 2, 3, 4, 52

[159] Y. Setty, A. E. Mayo, M. G. Surette, and U. Alon. Detailed map of a cis-regulatory input function. *Proc. Natl. Acad. Sci. USA*, 100:7702–7707, 2003. 5

[160] R. Shukuya and G. W. Schwert. Glutamic acid decarboxylase. I. Isolation procedures and properties of the enzyme. *J. Biol. Chem.*, 235:1649–1652, 1960. 18

[161] D. A. Siegele and J. C. Hu. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. USA*, 94:8168–8172, 1997. 7, 8, 21, 24, 37, 38, 42, 52

[162] J. L. Slonczewski, M. Fujisawa, M. Dopson, and T. A. Krulwich. Cytoplasmic pH measurement and homeostasis in bacteria and archaea. *Adv. Microb. Physiol.*, 55:1–79, 317, 2009. 11

[163] J. Snider, I. Gutsche, M. Lin, S. Baby, B. Cox, G. Butland, J. Greenblatt, A. Emili, and W. A. Houry. Formation of a distinctive complex between the inducible bacterial lysine decarboxylase and a novel AAA+ ATPase. *J. Biol. Chem.*, 281:1532–1546, 2006. 18

[164] W. Soksawatmaekhin, A. Kuraishi, K. Sakata, K. Kashiwagi, and K. Igarashi. Excretion and uptake of cadaverine by CadB and its physiological functions in *Escherichia coli*. *Mol. Microbiol.*, 51:1401–1412, 2004. 12

[165] V. Sourjik and H. C. Berg. Functional interactions between receptors in bacterial chemotaxis. *Nature*, 428:437–441, 2004. 5

[166] F. St-Pierre and D. Endy. Determination of cell fate selection during phage lambda infection. *Proc. Natl. Acad. Sci. USA*, 105:20705–10, 2008. 21

[167] D. M. Stoebel, A. M. Dean, and D. E. Dykhuizen. The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. *Genetics*, 178:1653–1660, 2008. 1, 2

[168] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456:516–519, 2008. 5

[169] S. Sundararaj, A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison, and D. S. Wishart. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of *Escherichia coli*. *Nucl. Acids Res.*, 32:D293–5, 2004. 4, 45, 46

[170] A. D. Tadmor and T. Tlusty. A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number. *PLoS Comp. Biol.*, 4: e1000038, 2008. 4

[171] S. Tănase-Nicola and P. R. ten Wolde. Regulatory control and the costs and benefits of biochemical noise. *PLoS Comput. Biol.*, 4:e1000125, 2008. 1, 5, 8, 52

[172] S.-W. Teng, Y. Wang, K. C. Tu, T. Long, P. Mehta, N. S. Wingreen, B. L. Bassler, and N. P. Ong. Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. *Biophys. J.*, 98:2024–31, 2010. 54

[173] L. Tetsch and K. Jung. How are signals transduced across the cytoplasmic membrane? transport proteins as transmitter of information. *Amino Acids*, 37:467–77, 2009. 9, 39

[174] L. Tetsch and K. Jung. The regulatory interplay between membrane-integrated sensors and transport proteins in bacteria. *Mol. Microbiol.*, 73:982–991, 2009.

[175] L. Tetsch, C. Koller, I. Haneburger, and K. Jung. The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Mol. Microbiol.*, 67:570–583, 2008. 12, 17

[176] L. Tetsch, C. Koller, A. Dönhöfer, and K. Jung. Detection and function of an intramolecular disulfide bond in the pH-responsive CadC of *Escherichia coli*. *BMC Microbiol.*, 11: 74, 2011. 9

[177] B. Teusink, M. C. Walsh, K. van Dam, and H. V. Westerhoff. The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.*, 23:162–9, 1998. 53

[178] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA*, 98:8614–8619, 2001. 42, 152

[179] M. Thattai and A. van Oudenaarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167:523–530, 2004. 7, 37

[180] J. M. Thevelein and S. Hohmann. Trehalose synthase: guard to the gate of glycolysis in yeast? *Trends Biochem. Sci.*, 20:3–10, 1995. 54

[181] D. M. Umulis, M. Serpe, M. B. O'Connor, and H. G. Othmer. Robust, bistable patterning of the dorsal surface of the drosophila embryo. *Proc. Natl. Acad. Sci. USA*, 103: 11613–8, 2006. 21

[182] J.-W. Veening, W. K. Smits, and O. P. Kuipers. Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.*, 62:193–210, 2008. 7, 54

[183] J.-W. Veening, E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. USA*, 105:4393–8, 2008. 6

[184] G. R. Venkatesh, F. C. Kembou Koungni, A. Paukner, T. Stratmann, B. Blissenbach, and K. Schnetz. BglJ-RcsB heterodimers relieve repression of the *Escherichia coli bgl* operon by H-NS. *J. Bacteriol.*, 192:6456–64, 2010. 27

[185] A. K. Vershon, S. M. Lioa, W. R. Mcclure, and R. T. Sauer. Bacteriophage P22 Mnt repressor. DNA-binding and effects on transcription *in vitro*. *J. Mol. Biol.*, 195:311–322, 1987. 152

[186] J. Vind, M. A. Sørensen, M. D. Rasmussen, and S. Pedersen. Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. *J. Mol. Biol.*, 231: 678–88, 1993. 1

[187] S. Wagner, M. L. Bader, D. Drew, and J.-W. de Gier. Rationalizing membrane protein overexpression. *Trends Biotechnol.*, 24:364–371, 2006. 4

[188] M. E. Wall, D. A. Markowitz, J. L. Rosner, and R. G. Martin. Model of transcriptional activation by MarA in *Escherichia coli*. *PLoS Comp. Biol.*, 5:e1000614, 2009. 16

[189] L. Wang, B. L. Walker, S. Iannaccone, D. Bhatt, P. J. Kennedy, and W. T. Tse. Bistable switches control memory and plasticity in cellular differentiation. *Proc. Natl. Acad. Sci. USA*, 106:6638–43, 2009. 34

[190] Y.-C. Wang and E. L. Ferguson. Spatial bistability of Dpp-receptor interactions during *Drosophila* dorsal-ventral patterning. *Nature*, 434:229–34, 2005. 21

[191] Z. Wang and J. Zhang. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc. Natl. Acad. Sci. USA*, 108:E67–E76, 2011. 8

[192] N. Watson, D. S. Dunyak, E. L. Rosey, J. L. Slonczewski, and E. R. Olson. Identification of elements involved in transcriptional regulation of the *Escherichia coli cad* operon by external pH. *J. Bacteriol.*, 174:530–540, 1992. 12

[193] J. C. Wilks and J. L. Slonczewski. pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J. Bacteriol.*, 189: 5601–5607, 2007. 10, 11

[194] D. M. Wolf, L. Fontaine-Bodin, I. Bischofs, G. Price, J. Keasling, and A. P. Arkin. Memory in microbes: quantifying history-dependent behavior in a bacterium. *PLoS ONE*, 3:e1700, 2008. 6, 21

[195] W. Xiong and J. E. Ferrell. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426:460–5, 2003. 24, 34

[196] Z. X. Yeo, S. T. Wong, S. N. V. Arjunan, V. Piras, M. Tomita, K. Selvarajoo, A. Giuliani, and M. Tsuchiya. Sequential logic model deciphers dynamic transcriptional control of gene expressions. *PLoS ONE*, 2:e776, 2007. 21

[197] A. Zaslaver, S. Kaplan, A. Bren, A. Jinich, A. Mayo, E. Dekel, U. Alon, and S. Itzkovitz. Invariant distribution of promoter activities in *Escherichia coli*. *PLoS Comp. Biol.*, 5: e1000545–, 2009. 1

[198] J. Zhan, B. Ding, R. Ma, X. Ma, X. Su, Y. Zhao, Z. Liu, J. Wu, and H. Liu. Develop reusable and combinable designs for transcriptional logic gates. *Mol. Syst. Biol.*, 6:388, 2010. 27

[199] Z. Zhang, W. Qian, and J. Zhang. Positive selection for elevated gene expression noise in yeast. *Mol. Syst. Biol.*, 5:299, 2009. 7

[200] D. Zilberstein, V. Agmon, S. Schuldiner, and E. Padan. *Escherichia coli* intracellular pH, membrane-potential, and cell-growth. *J. Bacteriol.*, 158:246–252, 1984. 10

Vielen Dank an ...

Prof. Dr. Ulrich Gerland, der mich mit seiner Faszination für die Biologie infiziert hat und mich immer wieder ermutigte tief in diese komplexe Welt einzutauchen. Dabei hatte ich alle Freiheiten die man sich wünschen kann, aber gleichzeitig einen absolut verlässlichen Mentor, der mir stets mit grossem Sachverstand und Intuition in stundenlangen Diskussionen zur Seite stand. Danke auch für Deine Geduld und die Möglichkeit dieses Promotionsprojekt mit den Bedürfnissen einer wachsenden Familie in Einklang zu bringen - auch dies habe ich sehr geschätzt.

Prof. Dr. Joachim Rädler, dessen vielseitiges Interesse an biophysikalischen Mechanismen stete Quelle der Inspiration darstellte und der die Einzelzell-Experimente zum Arabinose System möglich machte. Vielen Dank auch für die Unterstützung während der Anfangsphase meiner Promotion.

Prof. Dr. Kirsten Jung für die fruchtbare Kollaboration, die inspirierenden Diskussionen und die Möglichkeit ganz nah an den Experimenten zu arbeiten.

Dr. Judith Megerle für die tolle Zusammenarbeit, die entscheidend zum Gelingen dieser Arbeit beigetragen hat. Danke auch für Deinen langen Atem wenn ich mal wieder mit einer "guten Idee" ankomme!

Dr. Ina Haneburger, **Dr. Christiane Koller** und alle anderen **Kollaborationspartnern** für die hervorragenden Experimente, auf denen diese Arbeit fußt.

Patrick Hillenbrand und **Noreen Walker** die im Rahmen ihrer Diplomarbeiten und darüber hinaus meine Arbeit sehr bereichert haben.

Dr. Wolfram Moebius für die lustigen Stunden im Büro und beim Klettern.

Nico Geisel mit dem wissenschaftliche Diskussionen immer ganz besonders viel Spass machen.

Jan-Tim Kuhr, **Patrick Hillenbrand**, **Dr. Susanne Gebhard**, und **Fabienna Arends** fürs Korrekturlesen.

Prof. Dr. Erwin Frey und den ganzen **Lehrstuhl Frey** für die nette Arbeitsatmosphäre.

Prof. Dr. Thorsten Mascher und **meine neue Arbeitsgruppe** für die tolle Aufnahme und die Möglichkeit in Zukunft selbst Experimente zu machen.

Eva und Winni Werner für spontane Babysitter-Einsätze.

meine **Familie** für den Rückhalt, das Vertrauen in mich und die bedingungslose Unterstützung. Ohne meine **Eltern** und **Schwiegereltern** wäre diese Arbeit nicht möglich gewesen. Danke!

Neyla und **Lia** dafür, dass Ihr zu uns gekommen seid und uns täglich daran erinnert was wirklich wichtig ist. Mit Euch geht die Sonne auf!

meine liebe Frau **Kirstin**, mit der es nie langweilig wird und die eine großartige Bereicherung für mein Leben ist!

Ich versichere, diese Arbeit selbstständig angefertigt und dazu nur die im Literaturverzeichnis angegebenen Quellen benutzt zu haben.

München, den 12. Januar 2012